APPENDICES

MAKING AN IMPACT: A Preferred Framework and Indicators to
Measure Returns on Investment in Health Research

Panel on the Return on Investments in Health Research
January 2009

CAHS
ACSS

Canadian Academy of Health Sciences
Académie canadienne des sciences de la santé

Appendices

Making an Impact

A Preferred Framework and Indicators to Measure Returns on Investment in Health Research

Report of the Panel on Return on Investment in Health Research

Canadian Academy of Health Sciences/Académie canadienne des sciences de la santé

http://www.cahs-acss.ca/

ISBN: 978-0-9811589-1-4

Suggested citation: Panel on Return on Investment in Health Research, 2009. *Appendices - Making an Impact: A Preferred Framework and Indicators to Measure Returns on Investment in Health Research*, Canadian Academy of Health Sciences, Ottawa, ON, Canada

Additional copies of this report are available from the Canadian Academy of Health Sciences or through the Internet, **http://www.cahs-acss.ca/e/assessments/completedprojects.php**

For more information about the Canadian Academy of Health Sciences, visit the CAHS home page at: **http://www.cahs-acss.ca/**

Appendices

The following appendices present commissioned papers in areas where the report is not able to provide details, cover the background for the main report, and present the approach taken to the assessment process. The commissioned papers cover assessing the impacts of research in pillars II, III and IV but do not cover pillar I, since basic biomedical research is the area where most has been said on understanding the impacts of health research.

Contents

Appendix A: Commissioned Papers

- **Pillar II: Clinical Research**

- **Pillar III: Health Services Research**

- **Pillar IV: Population and Public Health Research**

- **Meso-Level Metrics for Impact**

- **Ethics and Evaluating Health Research**

- **Public Perspective on Health Research Funding**

- **Health Research Evaluation Frameworks: An International Comparison**

Pillar II: Clinical Research

How to Optimally Measure the Impact of Health Research Funding in Clinical Research

Ralph M. Meyer, M.D., FRCP(C)
National Cancer Institute of Canada Clinical Trials Group and Departments of Oncology, Medicine and Community Health and Epidemiology, Queen's University

Address correspondence to:
Dr. Ralph M. Meyer,
Director, National Cancer Institute of Canada Clinical Trials Group,
Cancer Research Institute,
Queens University,
10 Stuart St, Kingston, Ontario, Canada, K7L 3N6
Phone: (613) 533 - 6430 / Fax: (613) 533 – 6511
RMeyer@ctg.queensu.ca

Introduction

There is an increasing need to demonstrate the value of the investments Canadians allocate to health care research. While improvements in longevity and health-related quality of life of Canadians exemplify research successes, enhanced abilities to more explicitly evaluate the worth of research projects, programs and directions are needed in order to determine and justify the magnitude and methodology of allocating future investments. Statistics Canada estimates that in 2003-2004, funding of health care research in Canada totalled $5.08 billion; it is further estimated that during this time period, direct and indirect funding (e.g., matching funds) from federal government sources was almost $1.74 billion (Canadian Institutes of Health Research 2004). Given the magnitude of this funding and the competing needs for resources within the research community and across society, the Canadian Institutes of Health Research (CIHR) have developed a broad initiative to develop methodologies to evaluate research (Canadian Institutes of Health Research 2005). The goal of this report is contribute to this process by describing parameters for evaluating the worth of clinical research; the term 'payback' is used to describe the concept of value returned on research investment (Buxton & Hanney 1996).

There are outstanding examples of health-related and economic benefits that result from investments in biomedical and health sciences research. Commonly cited examples of research that have produced large returns on investment include the control of infectious diseases such as tuberculosis, polio, and influenza; treatment of hypertension and reductions in cardiovascular mortality and stroke; and the treatment of psychiatric disease including manic-depressive illness and schizophrenia (Silverstein et al 1995; Pardes 2000). These examples further demonstrate the need and value for breadth in the categories of research investment. Ultimate value to society has resulted from *basic* (or discovery) biomedical research. Observations from basic research are applied and tested in humans through

clinical research. Clinical and laboratory-based biomedical research now often include an iterative process that is part of *translational* research, in which there is linking of the study of biological specimens to clinical data. The application of clinical research to broader populations within societies can be evaluated through *health services* research, with intermediary steps between these research domains including the processes of *knowledge synthesis* and *knowledge transfer*.

This report will emphasize parameters associated with payback resulting from investment in clinical research. It is important to qualify this evaluation by indicating that basic, clinical and health services research form a continuum, including the basic-clinical interface of translational research and the clinical/heath-services interface through knowledge synthesis and transfer. Research across this continuum is required for ultimate large-scale societal benefits that include profound improvements in lives saved, quality of life and economic growth. Furthermore, preliminary and intermediary indicators of the benefits resulting from clinical research will include its influence and adoption into subsequent basic and health services research. While the interdependencies of these research domains are essential, the more specific benefits resulting from clinical research will be separately discussed in this report.

The Spectrum of Clinical Research

Clinical research encompasses a range of methodologies that have different objectives and utilities. At a broad level, clinical research includes the categories of descriptive and comparative research; comparative research can be further subdivided into research directed at identifying cause-effect relations and research that is directed at the quality of processes (Feinstein 1985). The objectives of these different forms of clinical research range from descriptive to generation of hypotheses to methodologies incorporating specific measures to control bias so that hypotheses can be formally tested. Descriptive and comparative research can be observational in nature and include *case series*, *case control* and *comparative cohort* designs. Clinical trials are a specific form of comparative research that is directed at evaluating cause-effect relations through the prospective evaluation of an intervention. Four types of clinical trials research exist: *phase I* trials typically are pharmacologically orientated and evaluate aspects of drug delivery such as dose and schedule; *phase II* trials are directed at determining estimates of efficacy and toxicities in cohorts receiving identical therapy; *phase III* trials are *randomized controlled trials* in which alternative interventions are compared in randomly assigned populations; and, *phase IV* trials are typically post-marketing/post-adoption surveys that prioritize identification of uncommon adverse events (Piantodosi 1997). Of these various forms of clinical research, only the phase III, or randomized controlled, clinical trial can reliably address the risks of bias and thus be used for hypothesis confirmation.

Implications for Research Payback

This report will assess research payback using the CIHR modification (Bernstein et al 2007) of a model developed by Buxton and Hanney (Hanney et al 2004). The term *'impact'* is used to describe and classify items of benefit resulting from research. Weiss, using reference to initiatives in the United States by the United Way, has contrasted use of the terms *'outputs'* and *'outcomes'* when describing these benefits (or impacts) and identifies three stages of potential outcomes (Weiss 2007). The initial stage addresses indicators of awareness of research conclusions; presentation and publication of results are common examples. An intermediate stage includes implementation of research findings; recommendations for, and demonstration of, adoption of research findings exemplify this stage. The final stage involves assessing the broader benefits resulting from adoption of research results, including parameters such as

improved health-related outcomes and economic benefits. These three steps are incorporated in greater detail using the CIHR framework, as described in the Methods of this report.

For each research initiative, crucial interactions exist between its methodology and conclusions and the subsequent adoption of these findings. In principle, conclusions of a research initiative, or at least interpretations of these conclusions, attempt to describe what might be considered as representing 'truth'. However, conclusions (or interpretations) may be misleading, through errors that are false-positive or false-negative. While interpretation and adoption of the conclusions of any research will affect all of the outcomes described by Weiss, adoption of findings from clinical research may have unique consequences, given the potential for their direct application to human subjects and patients. Therefore, special considerations are required when evaluating the worth of clinical research endeavours; dissemination and adoption are associated not only with potential benefits, but also with potential harms when false-positive or false-negative conclusions are applied to human populations. Confidence in the potential for benefit and cautions regarding the risks of harm will relate to research methodology. While research methodology may be accounted for during both the awareness and adoption processes (e.g. methodological quality is evaluated during peer-review and guideline/policy-determining processes), adoption of clinical interventions that have been developed using methodologies associated with a lower quality of evidence is well described (Guyatt et al 2006). Thus, payback through confirmed improvements in the broader and more mature outcomes related to population health and economic benefit will provide the highest quality evidence of research value, as these parameters are indicators of real, rather than potential, societal benefit.

Methods

The general objective of this review is to use the research payback model of Buxton and Hanney (Hanney et al 2004) to place into context the potential value of clinical research, including deliverables to society, that result from this investment. Specific objectives include:

1. To evaluate the potential application of the CIHR modification (Canadian Institutes of Health Research 2005) of the Buxton-Hanney model as a conceptual approach to consider payback from clinical research;
2. To review specific examples of attempts to evaluate payback from clinical research: these examples come from evaluations of geographic entities (e.g., research funding from governmental agencies) and specific populations (e.g., specific disease entities);
3. To examine the deliverables described in the examples evaluating payback from clinical research; and
4. To synthesize the findings from the above objectives into a context that is applicable to Canadians.

Between May 1 and July 4, 2008, two main methodologies were attempted. First, there was an initial attempt to perform a systematic review, with identification of eligible articles through a computerized literature search using MEDLINE. Search terms used included 'research' (including the specific categories of 'health services research', 'clinical research', 'biomedical research'), 'return on investment', 'research payback', 'research funding payback', 'health research deliverables', and 'economic return'. This strategy yielded few results. A second methodology was therefore utilized and incorporated the principles of an environmental scan. A search for relevant material was undertaken using the advanced search format of Google™; terms similar to those used in MEDLINE were used. This strategy yielded multiple relevant articles. The references and citations of these articles were then examined for additional relevant

citations. Articles published in English language periodicals were reviewed; in addition, several monographs produced by governmental research agencies and bodies and charitable research foundations were examined. Citations were excluded if they were evaluations of a specific research project (i.e., a single case study) or if they evaluated research payback related to a more narrow disease entity as opposed to a disease category; for instance evaluations of cardiovascular disease would be included whereas myocardial infarction was not. There was no attempt to review documents produced by or related specifically to for-profit organizations (e.g., the pharmaceutical industry). Furthermore, citations evaluating the benefits and/or limitations in the application of existing knowledge were not included.

Among the citations found using the above strategy was a critical review published in 2004 (Buxton & Hanney 2004); a similar search strategy was described in that manuscript. These authors also indicated that identification of reports of research payback was not straightforward as many of the reports were found in difficult to access books, monographs and reports with relatively few reports found in traditional journal publications. The intent of this current report is to address the specific objectives described above by identifying broad external trends, and specifically trends that might be more applicable to Canadian research, and to anticipate any challenges that might be encountered by Canadian funders of research in translating any resulting conclusions. Citations were evaluated for relevance and categorized as follows:

1. Citations describing overall context, methodology and relevance;
2. Citations analyzing geographic or specific funding agency performance;
3. Citations evaluating specific patient populations.

From these citations, an attempt was made to identify overlapping deliverables and concepts. From this inventory, priorities were established based on frequencies of response and an overall impression of importance. There was no independent review or consensus process; the conclusions reached are those of the author. Thus, the methodologies of article identification, retrieval, evaluation and interpretation place this manuscript into the category that is intermediate between that of a descriptive review and a formal systematic review.

The model of Buxton and Hanney (Hanney et al 2004) was applied as modified by the Canadian Institutes of Health Research (Canadian Institutes of Health Research 2005; Bernstein et al 2007). This model includes five core research outcomes (deliverables), which extend from specific to broad measures that include:

1. Knowledge Production;
2. Research Targeting and Capacity;
3. Informing Policy;
4. Health and Health Sector Benefits; and,
5. Economic Benefits.

The CIHR modification of this model collapses this schema into four categories with components of Buxton and Hanney's second category, Research Targeting and Capacity, allocated into the pre-existing first (Knowledge Production) and third (Informing Policy) categories (Bernstein et al 2007). The four new categories are therefore:

1. Advancing Knowledge
2. Informing Decision-Making
3. Health Impacts
4. Economic Impacts

Given the complexity of multi-national research and continuous changes in social policies and structure, the attribution of an outcome, and particularly a very broad outcome, to a specific research project is problematic and fraught with potential controversy. As the general objective of this review was to provide context, there was no attempt to validate the proportions by which specific research contributed to broad outcomes. Where possible, interpretations and conclusions were reached using a societal perspective.

Results

The CIHR has synthesized a list of potential indicators for evaluating research payback by category (Bernstein 2007), as shown in Table 1. In this section, evidence and commentary retrieved from the described search strategy will be used to augment CIHRs description of these indicators and presented in the Discussion section.

1. Citations Describing Overall Context, Methodology and Relevance
Knowledge Production (Awareness)
There is extensive literature in the science of bibliometrics assessing journal impact factors versus citation indices as surrogate indicators of research quality. Seglen (1997) summarizes some limitations of using journal impact factor as an indicator of research quality; more than twenty types of bias, measurement error and other limitations were listed. Major limitations relate to the lack of correlation between impact factor of a journal and the citation index of its individual articles and variation related to the field of research; publication in a journal with a high impact factor does not independently translate to more frequent citation. Seglen forcefully concludes that citation factors are a measure of 'scientific utility' rather than of quality and alternative measures are required for assessment rather than utilizing 'basically useless indicators'.

Druss (Druss & Marcus 2005) systematically reviewed the publication records of 18,211 R01 grants funded in 1996 by the US National Institutes of Health. These grants are 'to support a discrete, specified, circumscribed project' and are not intended to provide funding for training or infrastructure. Publications related to these grants were tracked over a nine year period; 199,009 citations were identified in 2943 journals with 22.2 percent being published in the Institute for Scientific Information's top 100 journals by impact factor. A mean of 7.58 manuscripts per grant were identified with a greater number of publications per grant observed for basic as compared with clinical research (8.39 versus 5.82; P < 0.001). The peak time to publication for grants dealing with basic research was four years with a gradual decline in the number of publications per grant thereafter. In contrast, publications related to clinical research peaked at five years and remained constant for the next two years before declining in year eight. Grants submitted for the category of competing renewal were associated with a greater number of publications than were new applications (7.43 versus 6.53; P < 0.001).

Weiss (2007) describes a conceptual approach to assessing research payback. Within his category of initial outcomes, in which awareness is assessed, he confirms the limitations of relying on journal impact

factor due to both its methodological limitations as a proxy indicator and furthermore suggests that eventual practice implementation is informed by more complex parameters such as formal and informal continuing medical education, interactions with colleagues and with non-medical personnel (e.g., from industry) as well as non-scientific publications found within the media and internet. He argues that more sophisticated measures that include 'sociometrics' (e.g., impact factor, citation index, non-journal publication) should be linked to better understandings of readership and that formal surveys of awareness would provide more accurate measures of penetration.

Informing Decision-Making (Adoption)

Grant et al (2000) evaluated United Kingdom practice guideline publications to assess the nature of the sources of the research that was incorporated into the guideline recommendations. Fifteen practice guidelines that were developed in the U.K. and approved by the National Health Services Appraisal Centre for Clinical Guidelines were evaluated; all were published between 1996 and 1998. The citations of these guidelines were evaluated by publication date, journal type and geographic site of research. In addition, a five percent sample of the citations from the guidelines were assessed to evaluate the citations included in these publications (i.e., a second generation analysis); further five percent samplings were performed of identified citations in order to complete third and fourth generation analyses.

For the fifteen guidelines, the median time between the date of the citation publication and the guideline publication was eight years; the peak time was six years and 25 percent of citations were published more than ten years prior to guideline publication. Among (what was then) the Group of Seven (G7) nations, most cited papers originated from the U.K. (25 percent) or the U.S. (32 percent); approximately seven percent originated from Canada. The types of journal in which citations were identified were classified as 'clinical observation' (e.g., *British Journal of Medicine*), 'clinical mixed' (e.g., *New England Journal of Medicine*), clinical investigative (e.g., *Immunology*) and 'basic' (e.g., *Nature*). The citations from the guidelines were predominantly found in clinical observation (approximately 30 percent) and clinical mixed (approximately 45 percent) journals; only 0.2 percent of citations were published in basic journals. As would be expected, with subsequent generations of citation analyses, the percentage of citations in clinical observation and clinical mixed journals decreased and citations from basic research increased; in a fourth generation analysis, eight percent of citations were published in basic journals.

At a more conceptual level, Weiss(2007) summarizes factors associated with the gap between awareness and adoption and lists five key factors that include the channel of communication, the integration of new findings with existing knowledge and practice (i.e., 'coherence'), the quantity and quality of the scientific evidence, parameters (such as issues related to feasibility) associated with the nature of potential change and the potential importance of implementation in terms of the magnitude of benefit and/or addressing an unmet need. He suggests that while initial processes to synthesize new knowledge, such as practice guidelines, are important, data from surveys regarding actual uptake, including use of electronic data bases to assess delivery practices across populations, are needed in order to provide data of the highest reliability.

Health Impacts and Economic Impacts (Implementation Leading to Benefit)

As indicated in the conceptual documents of Buxton and Hanney (1996) and in the CIHRs synthesis of research payback (Bernstein et al 2007), this topic relates to ultimately real (as opposed to proxy) evaluations of research worth. The methodology for measurement is far more complex. Health impacts

include the spectrum of disease prevention, detection/diagnosis, treatment and palliation. Benefits can be assessed using outcome measures that include potential years of life lost, quality-adjusted life years and disability-adjusted life years. Categories of economic benefit include new commercialization, direct savings and influence on human capital.

2. Citations Analyzing Geographic or Funding Agency Performance

The material retrieved from these reports can be more concisely described by report rather than by payback category.

Broad Initiatives

Australia

Kingwell et al (2006) reported survey results of paybacks of research from Australian National Health and Medical Research Council grants that completed funding in 1992, 1997 and 2003. Grantees were surveyed to provide information on publications, trainees, resulting changes (and benefits) in health care delivery and economic benefits. Survey responses were 61 of 139 (44 percent) in 1992, 131 of 259 (51 percent) in 1997 and 131 of 454 (29 percent) in 2003. Given these response rates, the reliability of the obtained data is limited. The mean number of publications per grant from funding that expired in 1997 was 7.0 for basic research and 5.2 for clinical research; similar data were obtained for 2003. Data regarding the number of trainees and patents per grant were also provided. Self reporting of health benefits included examples of change in practice and improved efficacy. The authors concluded that an interval of five to seven years from the time of expiration of funding to reporting should be considered in order to provide time for benefits to accrue while reducing the risk of compromised survey returns. Mechanisms to enhance survey comprehensiveness were described.

Hong Kong

Kwan et al (2007) evaluated payback of research funded by the Health and Health Services Research Fund of Hong Kong (HHSRF). A survey questionnaire, modified after the concepts of Buxton and Hanney, was pilot tested and then administered to 285 investigators who had completed projects with funding received funding from the HHSRF between 1993 and 2006. The questionnaire included six sections: knowledge production; use of research in the research system; use of research project findings in health system policy/decision making; application of the research findings through changed behaviour; factors influencing the utilization of research; and, health/health service/economic benefits. The response rate was 86.8 percent; higher response rates were associated with a shorter duration of time since expiration of funding. The mean duration of time since expiration of funding among respondents was 6.34 years.

The mean number of publications per grant was 5.4 with a higher publication rate noted for projects receiving greater funding amounts. Acquisition of higher academic qualifications (i.e., obtainment of academic degree) was indicated in 38.2 percent of projects, research directly led to subsequent research funding in 44.9 percent and projects were reported to have influenced treatment guidelines, protocols and standards in 35.4 percent. Factors associated with this latter deliverable included participation of the researcher on policy/advisory committees and an increased size of the funding award. The authors compared their results with those resulting from application of the Buxton and Hanney model in the U.K. and indicated that findings were similar.

Specific Initiatives

United Kingdom

Soper and Hanney (2007) describe a specific U.K. research funding initiative to address principles that underlie the adoption of findings from clinical research. The National Health Services Research and Development Implementation Methods Program was developed in 1994 to support research into components of knowledge syntheses and transfer. Twenty categories of research addressing principles such as source / method of research presentation and dissemination, roles and interventions of health care professionals / managers / administrators, feasibility and education were considered. The impact of the research funded by this initiative was evaluated using categories of payback developed by Buxton and Hanney. Summaries of publications, training, knowledge related to improving the capacities and success of such research and dissemination of findings and implications for knowledge transfer were summarized. Case studies were reported. While the author's main conclusions were that they had identified obstacles faced by those developing grant applications to address the specific issues of knowledge transfer and in developing broader strategic programmes of research in general, they also demonstrated that the model used to evaluate research payback was applicable.

Alberta

Buxton and Schneider (1999, 2008) described an evaluation of the Buxton and Hanney payback model as applied to research funded by the Alberta Heritage Foundation for Medical Research (AHFMR). Case study methodology was employed to evaluate four examples of funded research. The major objective of the evaluation was to determine if the model was generalizable to research, including clinical research, conducted in a geographic region that differs from where the model was developed. The justification for this objective included the fact that granting agencies, academic institutions and grantees have differing objectives, priorities and capacities that would affect assessment of research payback. For instance, prioritization of local needs may require modifications of methodologies used to assess a payback model that prioritized evaluation of research within a global environment. The authors concluded that the model was generalizable and suggested that utilizing the principles of the model might enhance the quality of grant submissions, evaluation and reporting of payback.

3. Citations Evaluating Specific Patient Populations.

Arthritis

In CIHRs evaluation of research payback, Buxton and Hanney are referenced with respect to a proposal to use their model to evaluate funding of arthritis research in the U.K. (Hanney et al 2004). The results from this evaluation were reported by Wooding et al (2005). The authors completed a comprehensive analysis of sixteen grants funded by the U.K. Arthritis Research Campaign; these grants were selected out of a pool of 556 in order to ensure that evaluation included a sufficient range of projects. Among the factors included in this selection process were project versus program applications, basic versus clinical research, investigator seniority, and background of the health professional. Each of the five categories of the Buxton-Hanney model was evaluated. Literature searches were performed to identify publications attributed to the investigators and citations of these were tracked. Systematic reviews and practice guidelines that referenced the research were obtained and when applicable, commentaries reporting the quality of the research were noted. Semi-structured interviews were conducted with the researchers. The research team of this evaluation project conducted a 2-day workshop to formally evaluate and reach consensus regarding the outputs.

The authors identified impacts across the range of output categories and concluded that the model for assessing payback was effective. Conclusions regarding research productivity included finding evidence across the spectrum of applications of impacts beyond peer-reviewed publications; even projects perceived to be of limited scope had diverse outputs and were assessed as providing value for money spent. Other conclusions were noting of the importance of the role of the investigators and their associated networks in leading downstream translation of research findings, the ability of investigators to capitalize on flexibilities in the granting system in order to forward research that ultimately provided value to the funder and the variable benefit of referee comments in the grant application process.

Neurologic Diseases

Johnston has reported the public return on investment of U.S. National Institutes of Health Neurological Disorders and Stroke (NINDS) funding of clinical trials by performing economic modeling that accounts for the costs of research and of subsequent changes in practice (Johnston et al 2006). All randomized controlled trials for which NINDS funding was completed by January 2000 were included. The authors reviewed the details and results of the trial intervention and performed economic evaluations in order to determine the cost per quality-adjusted life year (QALY) associated with conducting the trial. Systematic literature searches were performed to facilitate determining the costs of treatment and to perform economic evaluations comparing the adopted experimental treatment with previous standard therapy; QALYs were similarly determined. The role of the NINDS trial was proportionately reduced if other clinical trials reported similar conclusions. Return on investment was determined by assigning a value to the QALY based on the U.S. gross domestic product.

Out of 72 clinical trials funded between 1977–2004, 28 were eligible for inclusion; the NINDS expenditures on these trials were $335 million. Published results were available for 27 trials; the experimental intervention was superior to control arm therapy in fourteen, inferior in three and no difference was detected in ten. Additional treatment costs resulting from adoption of superior therapies was estimated to be $3.8 billion. The 10-year estimated benefit resulting from the research was 470,339 QALYs with an estimated cost-effectiveness ratio of $7,713 per QALY. Based on U.S. GDP, the net benefit of the projects was estimated to be $15.5 billion, representing a return on research investment of 4600 percent.

Cardiovascular Diseases

Clay has reported the results of a bibliometric analysis of research funded by the National Heart Foundation (NHF) of Australia (Clay et al 2006). Reports of cardiovascular research funded by the NHF that were published in 1996-2000 averaged 6.1 citations per publication versus 5.4 citations per publication in a control group of publications identified from similar literature. The authors commented that fewer NHF-funded projects may have been published in the specialty's highest impact journals.

Diabetes

Hanney has examined the impact of diabetic research using a novel process for tracking the results of one research program (Hanney et al 2006). The premise of this project was founded on the concept that markers of research impact can be determined in a 'forward' manner in which original research is tracked through subsequent citation bibliometrics, evaluations of the importance of the original research in contributing to the work in which it is cited and qualitative analyses that include completion of questionnaires by the original researchers, critical pathway analyses by experts in the field and interviews with the principle investigator of the original research. Based on productivity within a calendar year, 29 original papers were identified that were associated with 799 citations (second

generation papers), which, in turn led to identification of 12,891 third generation citations. The evaluation of the importance of the original work to the second generation papers was considered by an external reviewer to be of considerable or essential importance in nine percent of cases. The qualitative analyses uncovered information that would have otherwise been unidentified relating to the career directions of junior investigators/trainees involved in the original research, subsequent grant funding (and productivity) and development of patents. The authors concluded that their methodology was feasible and valid but also noted that the approach was labour-intensive.

Discussion

Given the broad scope, tremendous resource implications and critical importance of biomedical and health care research initiatives in Canada, it is vital that there be robust processes to measure resulting benefits. These processes must extend from assessments of the success associated with meeting specifically stated research objectives to the nature of broader societal benefits. Recognizing this need for rigorous evaluation, CIHR has initiated a process to systematically evaluate and categorize research paybacks (Bernstein et al 2007). With modifications, the CIHR process has relied heavily on the work of Buxton and Hanney (Buxton & Hanney 1996; Hanney et al 2004). The objective of this review is to use these payback models to, at a broad level, place into context the societal benefits of clinical research.

With this review, substantial evidence is provided that evaluates the validity and generalizability of the model of paybacks described by Buxton and Hanney. The model was found to be effective in describing payback benefits for research into different diseases including diabetes (Hanney et al 2006) and arthritis(Wooding et al 2005), research conducted in various geographic regions including the United Kingdom (Wooding et al 2005), Hong Kong (Kwan et al 2007) Australia (Kingwell et al 2006) and Alberta (Buxton & Schneider 1999, 2008), and across the spectrum of research categories ranging from more basic biomedical research (Hanney et al 2006) to health services research (Soper & Hanney 2007). Throughout these reports, there is evidence of validation and generalizability of using these models to evaluate clinical research.

It is not surprising that a system to evaluate payback associated with any form of biomedical or health care research would be applicable to clinical research as the indicators of the deliverables of successful research are largely generic. What might vary is the proportionate importance of various indicators and the manner by which these are evaluated and valued. An analysis of the potential for unique benefits (and risks) associated with clinical research may be facilitated by using an analogy that draws a parallel between the specific objectives, and thus deliverables, of the different forms of clinical trials research and the categories of deliverables described in the Buxton-Hanney and CIHR models. This analogy is based on the concept of endpoint categorization, which includes the use of surrogate and intermediate endpoints versus direct evaluations of major outcomes (Fleming 2005).

At its broadest level, the objectives of clinical research will normally be to improve the health of populations and to prevent or reduce the morbidity and mortality associated with disease processes. These objectives may be met through descriptions of diagnostic interventions including screening, evaluations of prognosis and testing of therapeutic interventions, including those that assess quality processes of health care delivery. Within the category of testing of new therapies, clinical trials research includes the systematic processes of phase I-IV evaluations (Piantadosi 1997); the outcomes associated with these trial types form a continuum in which the results from earlier phases of testing enable subsequent testing. For instance, determination of a safe drug dose and schedule in phase I testing permits phase II testing in which estimates of benefit and harm can be observed; these results will

determine whether phase III comparative testing is justified. Within phase III testing, trials may be of an explanatory nature and evaluate efficacy or alternatively be of a pragmatic nature in which effectiveness is assessed (Schwartz 1967). In this paradigm, pragmatic phase III trials provide the highest level of evidence for benefit (or harm); the more preliminary, or developmental, forms of clinical trials research often evaluate surrogate or intermediate endpoints that are used to enable processes to eventually conduct pragmatic phase III trials.

The categories of deliverables listed in the Buxton-Hanney and CIHR models could be viewed using a similar paradigm. The highest level of evidence of benefit from clinical research will come from the third and fourth categories of the CIHR model that respectively assess Health Impacts and Economic Impacts. For benefits in these categories to be realized, the enabling (or surrogate/intermediate) deliverables associated with the first and second CIHR categories of Advancing Knowledge and Informing Decision Making, which include what Weiss refers to as 'awareness' and 'adoption (2007) must first occur. While demonstrating payback indicators in Advancing Knowledge and Informing Decision Making is necessary for observing deliverables in the categories of Health Impacts and Economic Impacts, such demonstration is not sufficient for the higher level deliverables of payback in these latter two categories.

Use of the paradigm described above may be helpful because of the unique potential for the results of clinical research to be applied directly to human populations, and therefore risk subjecting these populations to the conclusions of research associated with false positive and false negative results. While there are substantial risks associated with misdirection of subsequent actions associated with erroneous findings of basic and translational research, direct dangers to the health of human populations are less likely. This paradigm and analogy can be exemplified by the results of the Women's Health Initiative (WHI) study evaluating hormone replacement therapy in postmenopausal women (Writing Group for the Women's Health Initiative Investigators 2002). This large, pragmatic randomized controlled trial substantially altered treatment recommendations for use of hormone replacement therapy by postmenopausal women as increased incidences of coronary heart disease, stroke, breast cancer and pulmonary embolism were observed with use of hormone replacement therapy. Previous treatment practices supporting this intervention were based on clinical research findings that came from more preliminary forms of clinical research, including non-interventional observational studies. Given the results of the WHI trial, these more preliminary forms of research would now be regarded as having provided false-positive results that were subsequently directly applied to postmenopausal women. Thus, were an evaluation of the payback of these earlier studies to have been performed prior to reporting of the WHI results, an overrating of merit would have been attributed based on assessments of the potential surrogate or intermediate categories of Advancing Knowledge (as judged by the bibliometrics of these studies) and Informing Decision Making (e.g., referencing and incorporation within a cautious but potentially supportive recommendation of a 2001 American Heart Association practice guideline (Mosca et al 2001)). In contrast, payback associated with the WHI study may ultimately achieve a higher level of benefit by providing deliverables in the more definitive payback category of Health Benefits.

A second potential unique form of payback from clinical research relates to deliverables that are appreciated at more local levels. Rationales and evidence for this potential come from several of the sources identified in this report. Given variations in disease incidence and prevalence associated with unique populations, funding agencies may develop their own special priorities in order to maximize the potential benefits of reducing disease burden in these populations. In addition, geographic priorities

may be influenced by the capacities associated with the need for personnel and technologic and financial resources that are required to address specific health-related problems. Were there to be local priorities that influenced the decision to support a clinical research initiative, it is rationale to conclude that evaluations of payback should include mechanisms to assess local indicators. Accounting for local factors in developing priorities was exemplified in the analysis of payback from AHFMR funding of research in Alberta (Buxton & Schneider 1999, 2008).

The results of this review suggest that deliverables from conducting research do have local implications. Weiss describes the complexity of processes associated with 'awareness' (i.e., advancing knowledge) and the important role of interactions with colleagues (Weiss 2007). In an evaluation of research payback in Hong Kong, Kwan noted that survey results from HHSRF-funded investigators showed that 35.4 percent of research projects were reported to have subsequently influenced treatment practices and that such adoption was associated with the roles of funded investigators on policy and advisory committees (Kwan et al 2007). Similarly, Wooding describes the importance of U.K. Arthritis Research Campaign-funded researchers in leading to the downstream translation of their own findings (Wooding et al 2005). Finally, Grant found that among practice guidelines produced in the U.K., 25 percent of citations used to support recommendations came from U.K-related research reports (Grant et al 2000). These findings are consistent with a thoughtful description of dissemination of health care practices provided by Berwick in which he describes characteristics of 'innovators' and 'early adopters' of health care practices and that:

> 'Medical communities are primarily local in their orientation, are dominated numerically by early and late majority groups, and do not trust remote and personally unfamiliar sources of authority. The counterweight ought to be a formal, deliberate, organized system of search for innovations' (Berwick 2003).

In addition to the influences clinical research may have in the local adoption of findings, the role of specific funding in the career development of a researcher in becoming a potential opinion leader and contributor/leader of institutional, provincial, national and international policy and advisory panels is a potential deliverable to be evaluated under the 'building capacity' concept of the Advancing Knowledge deliverable.

This review also uncovered potential methodological implications for assessing payback associated with clinical as opposed to non-clinical research. With respect to Advancing Knowledge, limitations and implications for the use of bibliometrics included the limitations associated with using journal impact factor (Seglen 1997), findings that clinical as compared with basic research was associated with a lower average number of publications (Druss & Marcus 2005; Kingwell, Anderson & Duckett 2006) and that greater durations of time were required to observe the peak rates of publication (Druss & Marcus 2005). With respect to Informing Decision Making, evaluations of incorporating research findings into practice guidelines may be an effective tool; however, sufficient time from the initial reporting of research findings is required as in one study the median time from this reporting until guideline publication was eight years (Grant et al 2000).

Several reports describing evaluations of research payback utilized surveys of and interviews with researchers in order to evaluate specific deliverables. Benefits and limitations were observed. Among the benefits were discovery of unique roles of researchers in downstream translation and adoption (Kwan et al 2007; Wooding et al 2005), the career-related benefits to junior investigators and trainees, including the role of the evaluated research in influencing subsequent grant funding obtained by these

individuals, and unique perspectives of the eventual importance of the research (Hanney et al 2006). Limitations included survey response rates and the potential labour-intensive nature of detailed interviews (Kingwell, Anderson & Duckett 2006; Hanney et al 2006). None of the articles reported specific evaluations of the risk of bias associated with these methodologies.

Surprisingly, the search strategy for this review did not uncover reports assessing use of data bases of health care utilization or outcome as tools to assess research payback. Such evaluations would seem ideal mechanisms to assess adoption of research findings as a marker of Informing Decision Making and benefits arising from this adoption under the Health and Economic Impacts categories. These data bases exist at institutional, provincial and national levels and could evaluate parameters such as drug utilization, surgical techniques, physician behaviours as assessed through billing codes, hospital-related activities as well as incidence and mortality data. Evaluations of these data bases could be descriptive in nature or take the form of hypothesis-based health services research; subsequent funding of health services research would then be an indicator of clinical research payback under the Advancing Knowledge category. Using either descriptive or hypothesis-based analyses, use of data bases would facilitate evaluations of the payback from clinical research in the categories of Health Impacts and Economic Impacts.

The Economic Impacts category is likely to be comprised of two major forms of evaluation of clinical research indicators. As described by Johnston, economic evaluations including cost-utility and cost-benefit analyses may provide both direct measures of the benefit resulting from clinical research and can be modeled to estimate returns on research investment (Johnston et al 2006). In addition, the expanding role of translational research, including the development of biomarkers, will become a crucial aspect of determining what may be referred to as 'personalized' health care practices (Morrow & de Lemos 2007) and will necessitate systematic evaluation through clinical trials testing specifically designed to evaluate biomarker utility (Sargent et al 2005). In addition to improvements in health care outcomes, the discovery and evaluation of biomarkers will be associated with intellectual property that forms the basis of patency.

The general objective of this review was to use the research payback model of Buxton and Hanney (Hanney et al 2004) to place into context the potential value of clinical research. The review has demonstrated that the topic of research payback is regarded with increasing importance with systematic initiatives for evaluation already established by governmental agencies in the U.K and Australia and by a charitable organization in the United States. Commonalities among these processes are evaluations of the dissemination and adoption of research findings, assessments of the health-related and economic benefits resulting from adoption and ensuring that funding initiatives contribute to sustaining a robust future research enterprise.

There were four specific objectives of this review. The first was to evaluate the potential application of the CIHR modification (Bernstein 2007) of the Buxton-Hanney model as a conceptual approach to consider payback from clinical research. As would be expected, given that the Buxton-Hanney model provides an effective mechanism for evaluating the breadth of biomedical and health care research, the model with the CIHR modifications appears valid in assessing clinical research. The second objective was to review specific examples of attempts to evaluate payback from clinical research; these examples were extracted from overall reviews of payback that included evaluations of different diseases and multiple funding agencies or initiatives that were based in various geographic settings. High degrees of overlap in the processes and findings from these reports were observed.

The third objective was to examine the deliverables described in the examples evaluating payback from clinical research. The evaluation of this objective suggested several conclusions. First, processes to assess clinical research are strongly aligned with overall process to evaluate all types of biomedical and health-related research as opposed to requiring special evaluation processes for clinical research. Second, unique interpretation of some indicators used to assess clinical research is appropriate: specifically, attention is required to the surrogate or intermediate nature of deliverables categorized under Advancing Knowledge and Informing Decision Making as compared with the more definitive categories of Health Impacts and Economic Impacts, given the risk of applying false-positive or false-negative results directly to human populations; and, clinical research may have greater 'local' implications related to eventual health care policies that are applied specifically within the sphere of influence of the researcher and potentially the funding agency. Third, while inclusion of some indicators may be highly generalizable to all forms of research, the methodology by which these are applied and weighted may require modifications when assessing clinical research.

Finally, the fourth objective was to synthesize the findings from the above objectives into a context that is applicable to Canadians. This objective has already been largely addressed by CIHR as is indicated in Table 1; with minor modifications to accommodate the specifics of the goals of other Canadian funders of research, these indicators are generalizable to these agencies. Special opportunities may exist within this model for the evaluation of clinical research in Canada; these may specifically relate to Canada's universal health care policies and the manner in which these are operationalized at provincial levels as these processes are associated with data bases that may both facilitate evaluation of deliverables in the more definitive Health Impacts and Economic Impacts categories and permit assessments of 'local' relevance.

Table 1: CIHR Determined Categorization of Research Payback Including Potential Indicators and Sources (Bernstein et al 2007)

Advancing Knowledge

Indicators

1. Number of discoveries/breakthroughs resulting from CIHR-supported research
2. Number of Canadian health research publications
3. Number of publications resulting from CIHR-supported research
4. Impact of publications as demonstrated by citation intensity (citations/GDP) compared with wealth intensity (GDP/population)
5. Percentage of Canada Research Chair (CRC) holders attracted to or retained in Canada
6. Number and type of PhD graduates in Canada by year
7. Percentage of PhD graduates in Canada planning post-doctoral work in health

Sources

Bibliometric studies
End of grant research results reporting
Citation impact analysis
Databases of CRC holders
Data available through Statistics Canada (i.e. census and survey data)
CIHR performance management data

Informing Decision Making

Indicators

1. Public policies informed by CIHR and CIHR-funded research
2. Clinical practice informed by CIHR-funded research
3. Health system management decisions informed by CIHR-funded research
4. Research, policy and/or practice agendas influenced by funded
5. Research, policy and/or practice agendas influenced by funded research and/or CIHR institutes
6. Impact of Canadian health research publications
7. Impact of publications resulting from CIHR-supported research

Sources

Case studies (multi-method special studies)
End of grant research results reporting
CIHR program evaluations
CIHR performance management data
Research user surveys
Citation impact analysis

Table 1 (continued)

Health Impacts
Indicators

1. Research study participants' health status affected by participating in CIHR-funded research
2. Population health status influenced by CIHR –funded research
3. Potential years of life lost (PYLL) for target disease categories (e.g. cancer, circulatory disease) influenced by CIHR-funded research
4. Health-related quality of life influenced by CIHR-funded research

Sources
Case studies (multi-method special studies)
End of grant research results reporting
Statistics Canada data
Special studies to establish links to health research
CIHR performance management data
Analyses of publications

Economic Impacts
Indicators

1. Number and nature of patients, spinoff companies and IP licenses influenced by CIHR-funded research
2. Income from IP commercialization
3. Commercial use of research funded by CIHRs commercialization programs
4. Cost savings influenced by CIHR-funded research
5. Human capital gains, including productivity influenced by CIHR-funded research

Sources
End of grant research results reporting
Statistics Canada data
Case studies (multi-method special studies)
Technology assessment special studies
Collaborative studies with Health Canada and Statistics Canada

References

Bernstein, A., et al., 2007. A framework to measure the impacts of investment in health research. Science, technology and innovation indicators in a changing world: responding to policy needs. *Organisation for Economic Cooperation and Development*, p. 231 – 250.

Berwick, D.M., 2003. Disseminating innovations in health care. *The Journal of the American Medical Association,* 289 (15), p. 1969 – 1975.

Buxton, M., Hanney, S., 1996. How can payback from health services research be assessed? *Journal of Health Services Research Policy*, 1 (1), p. 35 – 43.

Buxton, M., Hanney, S., 2004. Estimating the economic value to societies of the impact of health research: a critical review. *Bulletin of the World Health Organization,* 82, p. 733 – 739.

Buxton, M.J., Schneider, W.L., 1999, 2008. Assessing the payback from AHFMR-funded research. *Alberta Heritage Foundation for Medical Research.*

Canadian Institutes of Health Research (CIHR), 2004. *Finding a balance in federal health research funding* (Health research results and related reports) [Online]. (Updated 26 April 2007) Available at: http://www.cihr-irsc.gc.ca/e/26467.html [accessed 10 July, 2008]

Canadian Institutes of Health Research (CIHR), 2005. *Developing a CIHR framework to measure the impact of health research.* (Synthesis report of meetings, February 23, 24 and May 18, 2005) [Online]. (Updated 20 Jan. 2006) Available at: http://www.cihr-irsc.gc.ca/e/30324.html [accessed 10 July, 2008]

Clay, M.A., et al., 2006. The returns from cardiovascular research: the impact of the National Heart Foundation of Australia investment. *Medical Journal of Australia*, 185 (4), p. 209 -212.

Druss, B.G., Marcus, S.C., 2005. Tracking publication outcomes of National Institutes of Health grants. *The American Journal of Medicine,* 118 (6), p. 658 – 663.

Feinstein, A., 1985. *A classification of medical research*. *Clinical epidemiology: the architecture of clinical research.* Philadelphia, London, Toronto, Mexico City, Rio de Janeiro, Sydney, Tokyo: W.B. Saunders Company.

Fleming, T.R., 2005. Surrogate endpoints and FDA's accelerated approval process. *Health Affairs,* 24 (1), p. 67 – 78.

Grant, J., et al., 2000. Evaluating 'payback' on biomedical research from papers cited in clinical guidelines: applied bibliometric study. *The British Medical Journal,* 320 (7242), p. 1107 – 1111.

Guyatt, G., et al., 2006. Grading strength of recommendations and quality of evidence in clinical guidelines: report from an American College of Chest Physicians Task Force. *Chest*, 129 (1), p. 174 – 181.

Hanney, S., et al., 2004. Proposed methods for reviewing the outcomes of health research: the impact of funding by the UK's 'Arthritis Research Campaign'. *Health Research Policy and Systems*, 2 (1), p. 4.

Hanney, S.R., et al., 2006. Identifying the impact of diabetes research. *Diabetic Medicine,* 23 (2), p. 176 – 184.

Johnston, S.C., et al., 2006. Effect of a US National Institutes of Health programme of clinical trials on public health and costs. *The Lancet,* 367 (9519), p. 1319 – 1327.

Kingwell, B., Anderson, G., & Duckett, S., 2006. Evaluation of NHMRC funded research completed in 1992, 1997 and 2003: gains in knowledge, health and wealth. *Medical Journal of Australia,* 184, p. 282 – 286.

Kwan, P., et al., 2007. A systematic evaluation of payback of publicly funded health and health services research in Hong Kong. *BMC Health Services Research,* 7 (1), p. 121.

Morrow, D.A., de Lemos, J.A., 2007. Benchmarks for the assessment of novel cardiovascular biomarkers. *Circulation,* 115 (8), p. 949 – 952.

Mosca, L., et al., 2001. Hormone replacement therapy and cardiovascular disease: a statement for healthcare professionals from the American Heart Association. *Circulation*, 104 (4), p. 499 – 503.

Pardes, H., 2000. The perilous state of academic medicine. *The Journal of the American Medical Association*, 283 (18), p. 2427 – 2429.

Piantadosi, S., 1997. *Clinical trials as experimental designs. Clinical trials: a methodologic perspective.* New York, Chichester, Weinham, Brisbane, Singapore, Toronto: John Wiley and Sons, Inc.

Sargent, D.J., et al., 2005. Clinical trial designs for predictive marker validation in cancer treatment trials. *Journal of Clinical Oncology,* 23 (9), p. 2020 – 2027.

Schwartz, D., Lellouch, J., 1967. Explanatory and pragmatic attitudes in therapeutical trials. *The Journal of Chronic Diseases*, 20 (8), p. 637 – 648.

Seglen, P.O., 1997. Why the impact factor of journals should not be used for evaluating research. *The British Medical Journal*, 314, p. 498 – 502.

Silverstein, S.C., Garrison, H.H., & Heining, S.J., 1995. A few basic economic facts about research in the medical and related life sciences. *The Journal of the Federation of American Societies for Experimental Biology*, 9 (10), p. 833 – 840.

Soper, B., Hanney, S., 2007. Lessons from the evaluation of the UK's NHS R&D Implementation Methods Programme. *Implementation Science*, 2 (1), p. 7.

Weiss, A.P., 2007. Measuring the impact of medical research: moving from outputs to outcomes. *The American Journal of Psychiatry,* 164 (2), p. 206 – 214.

Wooding, S., et al., 2005. Payback arising from research funding: evaluation of the Arthritis Research Campaign. *Rheumatology*, 44 (9), p. 1145 – 1156.

Writing Group for the Women's Health Initiative Investigators, 2002. Risks and benefits of estrogen plus progestin in health postmenopausal women: principal results from the Women's Health Initiative Randomized Control Trial. *The Journal of the American Medical Association,* 288 (3), p. 321 – 333.

Pillar III: Health Services Research

Estimating the Return on Investment from Health Services Research: A Theoretical and Empirical Analysis

Steven Lewis

Patricia J. Martens PhD
Director, Manitoba Centre for Health Policy Associate Professor, Department of Community Health Sciences, Faculty of Medicine, U. of Manitoba; CIHR/PHAC Applied Public Health Chair.

Louis Barre

October 2008

I. Introduction

Health services research (HSR) is a diverse enterprise. The Agency for Healthcare Research and Quality definition is as good as any:

> Health services research examines how people get access to health care, how much care costs, and what happens to patients as a result of this care. The main goals of health services research are to identify the most effective ways to organize, manage, finance, and deliver high quality care; reduce medical errors; and improve patient safety (2002).

These goals imply utility: understanding the effectiveness and efficiency of health care should result in improvements. But some HSR is curiosity-based and often produces insights as fundamental and original as the fruits of basic science studies.

HSR, like all health research, is expanding. The Canadian Institutes of Health Research (CIHR) was launched in 1999 as part of an overall policy goal of catapulting Canada into the top 5 research countries in the world. Increased funding begets higher expectations and more intense scrutiny. This is especially so for applied research. Much HSR is commissioned to address particular issues. The pathway from the findings to the impact is expected to be shorter and more direct. One could conceive of a general equation for expected research funding as:

$$\$ \sim (m,f,h,e), \text{ where}$$

- m=the perceived *magnitude* of the problem or challenge it is designed to address

- f=the *fear* factor for the problem to be solved
- h=the degree of *hope* the potential solution inspires
- e=the amount of *excitement* that a possible breakthrough engenders

The coefficients for all 4 variables have in the past been seen as low for HSR. Demonstrating that some hospital stays are needlessly long is not quite like landing on the moon. HSR debunks at least as often as it affirms, and the prospect of reducing unnecessary laboratory tests or evidence of diminishing returns from additional MRI scanning can, to some funders, seem less glamorous than some basic science premises such as stem cell research.

Not all HSR responds to externally articulated demand. Some is conceived by researchers in much the same way that investigators conceive basic biological studies. Even many commissioned studies aim to generate fundamental new insights and begin discussions rather than prescriptions for action. One study of commissioned research found that 40% of completed projects directly influenced policy (Innvaer et al 2007)[1]. Many evidence-based products specifically designed to affect day-to-day health care practice – such as clinical practice guidelines - often have little impact (Cabana et al 1999). Whether research is applied depends on its quality, timeliness, and accessibility – elements for which the research process and structure might reasonably be held accountable - but even more on the decision-making culture and environment in which it takes place.

For these and related reasons, despite major advances in research expertise and methodological sophistication in recent decades, estimating the return on investment (ROI) from HSR and medical research remains an inexact science. Leading international researchers summarize the current state of the art as follows:

> Overall, there is a growing evidence base demonstrating that health and biomedical research is an investment: there are tangible benefits and it is quite possible that exceptionally attractive long-term returns may accrue. Substantial efforts are, however, needed to refine existing methodologies and to make them more robust if we are to move from suggestive studies to firm estimates that cannot easily, as now, be challenged and contested (Buxton, Hanney & Jones 2004, p. 734).

II. Measuring Impact: Conceptual and Practical Challenges

The problems with establishing a cause-and-effect relationship between HSR[2] and concrete outcomes have been well described. Chief among them are:

1. Research is only one potential influence on decision-making. While there is near-unanimous agreement that research-based evidence should be more influential in decision-making, values, beliefs, preferences, traditions and culture, politics, considerations of fairness, entitlements, and interests remain powerful. Our society is not a research-ruled technocracy; it is a democracy, where decisions must be seen as not only technically sound, but also legitimate, which has connotations of fairness, responsiveness, representativeness, and consistency with embedded norms and preferences. Various interests and constituencies influence whether and how research is integrated into policy and practice (Hanney et al 2003).

---

[1] That figure may be optimistic given that it is based on policy-makers' self-reported assessment rather than any objectively traced connection between findings and policy.

[2] Many of these issues could also apply to other pillars of research.

2. Concrete action is the last stage of a process that begins with awareness, and completing the journey from awareness of evidence to its widespread implementation may take a long time. It takes up to 17 years for the findings of high-quality randomized controlled trials to become standard clinical practice (Balas & Boren 2000; Institute of Medicine 2001). Canadian knowledge translation (KT) experts refer to a 'ladder of knowledge utilization' with six rungs: transmission, cognition, reference, effort, influence, and application (Landry, cited in Hanney et al 2003). Tracing the final result to its research origins after a major time lag may be difficult.

3. Arguably, the greatest impact may also be less tangible and quantifiable - when research creates new awareness, understanding, and discussion that contribute to cultural change. Original insights and ideas do not become concrete actions overnight. The more truly innovative the finding, the longer it is likely to take to affect policy and practice. A new way of looking at the world – e.g., the population health perspective built on research linking non-medical determinants of health to risk factors, gradients in service use and outcomes, etc. – can affect individual practice and public policy in countless ways, but tracing the pathway from insight to action is difficult.

4. Research is rarely stand-alone; it takes place in a context, it usually builds on and becomes part of a body of related work, and where it is commissioned, it is possible that decision-makers are already primed to act. Moreover, decision-makers may be influenced by a host of research and non-research findings, making it difficult to attribute impact accurately to any single source. Correlation is not necessarily causation.

5. The path to policy development and resource allocation is rarely short, and often research confirms or justifies rather than uniquely informs what decision-makers intend to do. Commissioning and carrying out research takes time, which by definition means that ideally, the deliverable agenda should anticipate the future decision-making environment rather than respond only to current issues. The future decision-making agenda is never entirely predictable. Quite possibly, findings that appear to have translated rapidly into policy and practice may in fact have been the serendipitous accompaniment to decisions that were already imminent. Distinguishing correlation from causation is difficult.

6. More often than not, it is impossible to define exactly what a 'perfect' response to research would be. Without consensus on what the ideal form of research implementation would be, it is impossible to estimate the gap between observed performance and the gold standard.

7. Research organizations have limited control over the ultimate use of their work. As a result, Hovland has proposed that the evaluation of impact focus more on changes in behaviours and relationships among those with whom they work than on concrete outcomes, and on contributions to, rather than sole responsibility for, the ultimate impact (Hovland 2007).

8. HSR is rarely unambiguously conclusive for every context, and often there are alternative interpretations of the same data. In many cases the prudent response of even a highly developed evidence-oriented culture is caution. Syntheses of similar studies are more difficult in HSR than in, say, clinical trials research because of uncontrollable or dissimilar contextual factors that may influence methods and results.

These complexities suggest a nuanced understanding of ROI. Impact and ROI are multi-dimensional, not limited to only on those that are easily measurable, visible within a short period of time, and translatable into cash equivalents.

III. Estimating ROI from the Manitoba Centre for Health Policy: A Case Study

Using a case study of how to estimate ROI for health services research provides a real life example of the complexities and issues associated with identifying returns. The Manitoba Centre for Health Policy (MCHP) is used as an example since it represents the authors best understood case study.[3]

The MCHP was established in 1991 as a university-based health services and population health research centre, grant-funded partially through ongoing 5-year funding grants from Manitoba Health and Healthy Living. The MCHP is expected to:

1. Produce 5 research projects (called 'deliverables') per year, topics to be mutually decided upon by the Ministry of Health & Healthy Living (MHHL) and the Director of MCHP;
2. Ensure knowledge translation of these deliverables (a 6th 'deliverable' is called a KT deliverable, to ensure funding for this process); and
3. Maintain and expand the Population Health Research Data Repository (referred to as the Repository) of linkable datasets, both for MCHP/Manitoba Health and Healthy Living deliverable purposes and for other research purposes.

Since 1997 the Centre has produced about 60 deliverable reports on a great range of topics. They can be categorized as:
1. Analyses of important aspects of the broader health system, e.g., studies of rural hospitals, wait times, personal care home ('nursing' home) quality of care, costs of various services.
2. Analyses of specific use of services among sub-populations, e.g., high users of drugs, end-of-life patterns of care, children's health.
3. Reports that increase understanding of health and health care issues, e.g., mental health, high school completion rates, health status and health care utilization atlases.
4. Reports that support planning and forecasting, e.g., acute care and personal care home bed needs, observation units in hospitals.
5. Methodological reports, e.g., case costing, needs-based funding, measurement of morbidity, primary care productivity and distribution.

Recently the Centre and Manitoba Health and Healthy Living commissioned a review to estimate the impact of the deliverables, including ROI. The Centre continues to be funded and its budget has been recently increased, suggesting that the government considers it a success. This implicit affirmative judgment is not based on precisely articulated and accurately measured performance indicators[4].

The MCHP experience is a near-ideal case study because circumstances here would seem conducive to generating a ROI from HSR. The government established the MCHP, demonstrating its commitment to the use of research to inform and evaluate both decisions and components of the system. It takes an active and ongoing interest in setting the deliverables agenda. There is continuous liaison between the

---

[3] It should be noted that this case study does not imply that the MCHP should be seen as the benchmark for other HSR organizations, rather that it identifies how such an evaluation might be conducted and the sorts of findings that might arise.

[4] Governments may also be somewhat reluctant to be too precise in their expectations because as commissioners of research and decision-makers, they largely determine whether research findings will be applied to policy and practice. An evaluation of the ROI in the Centre is at least partly a self-evaluation of government.

MCHP and the health ministry. To varying degrees ministry officials participate in project working groups.

The pathway from conception of a deliverable to its eventual impact varies considerably. A deliverable to estimate the potential impact of reference-based pricing lends itself to a yes-no choice as to whether to implement the policy. A deliverable that describes variation in prescribing patterns and costs has implications for clinical practice; improved performance requires commitment and behavioural change at the policy, organizational and individual level. In Canada's highly autonomous practice environment, impact will be variable and difficult to trace.

Taking these factors into consideration, consistent with the literature on research impact, and based on a review of the deliverables produced by the Centre, this case study attempted to identify ROI and impact in the following areas:

Policy and the Culture of Decision-Making

- Influencing the policy agenda by bringing new findings and insights to the attention of decision-makers
- Policy development – influencing legislation, regulation, negotiations, directives, overall budget priorities, target populations
- Policy implementation – establishing or modifying programs and services, designing the structures and rules
- Changing the nature of policy discussions by bringing new information to the table.

Financial

- Cost savings in a particular area (e.g., a reduction in utilization or price)
- Capital cost avoidance (e.g., a decision not to build new beds)
- Improved cost-effectiveness of services (e.g., greater productivity, reduction of waste, more service for the same resources)
- Distribution of resources (e.g., better alignment of needs and resources, enhanced fairness of system, identification of unmet needs)
- Research grants received by Centre staff and/or by others for whom the Centre infrastructure and Repository were central components
- Personnel awards won by researchers whose work is significantly dependent on the Centre data and infrastructure.

Health Status

- Effectiveness of services
- Longer term health status outcomes among population groups
- Identification of new health status issues or dimensions

Public Confidence

- Concrete and objective information on access, quality, and outcomes that balances anecdotal, biased, or incomplete sources that affect public opinion
- Evidence that problems are being addressed (time trends, etc.)

- Evidence that a reported problem or deficiency is overstated.

Capacity Building

- Increased awareness of and receptivity to research-based evidence in government, RHAs and other health organizations
- Training and support for analysts, researchers, planners and decision-makers in RHAs
- Role of the Repository in enhancing health services and population health research in the province
- Academic personnel recruited to and/or retained in Manitoba for whom the Centre infrastructure is a key element of their careers.

Central to the challenge of quantifying impact and ROI is that there is no formal articulation of how deliverables might affect policy, practice, and resource allocation. There are no commitments to act on findings or achieve quantifiable targets. Some projects aim to generate fundamental new knowledge rather than immediately applicable policy-oriented information.

That said, both the Centre and Manitoba Health and Healthy Living work to ensure that the findings reach an array of decision-makers. Within the ministry, an Assistant Deputy Minister (ADM) and a Director sponsor each deliverable, and provide general oversight, consultation and advice as needed while projects are in progress. The MHHL does not micro-manage projects and recognizes that the integrity of the final products depends on adherence to high academic and methodological standards. They or their staff may also be members of the deliverable working group. The ADM and Director brief the Ministry on draft report highlights and implications, and actions to be taken or mandated as a result.

MCHP researchers and support staff also pursue a diversified dissemination and knowledge translation strategy. They engage in numerous briefings, periodic policy discussions, and broad-based dissemination through annual workshops with MHHL, the Winnipeg Regional Health Authority (WRHA), and the 10 health regions outside Winnipeg. These dissemination and engagement efforts are unusually intensive and sustained. This 'KT intensity' is attributable to the sense of ownership on the part of Manitoba Health and Healthy Living for the deliverables, as well as the Centre's own leadership, commitment to, and expertise in presenting the findings to a range of audiences and developing a research-friendly health system culture.

IV. Methods, Limitations, and Approach

While there are unresolved conceptual issues in estimating HSR ROI, there is an even more important barrier: the absence of comprehensive and reliable data. As a result, we pursued a number of methods to assess impact and ROI. These are:

1. Interviews with the principal audiences of MCHP deliverables – Manitoba Health and Healthy Living and the RHAs. Interviewees were also asked to estimate, where appropriate and feasible, the quantitative impact of projects. The qualitative approach dominates the literature. It has the advantage of examining impact from a number of perspectives and creates a connective story around the context the findings are intended to inform. The limitation is that there is no way to verify which account of reality is correct, especially when there are differing perspectives on what became of a deliverable. Some may consider anything short of hard, quantifiable data as little more than informed opinion highly subject to inaccuracy and/or bias. Additionally, there

may be an inclination toward a 'socially desirable response' that overstates the role of research given the contemporary cultural status of evidence-informed decision-making.

2. First-hand reports of decisions that might reasonably be attributed significantly to a deliverable. Some deliverables have dealt with concrete issues, such as the substitutability of drugs, or the adequacy of bed numbers, that can lead to clear-cut decisions and actions. An example is bed forecasting: did the government accept the Centre projections; did they affect RHA planning and priorities; were capital projects pursued, deferred, or shelved, etc.

3. Examinations of utilization and cost trends before and after relevant projects were undertaken. The pharmaceutical deliverables lend themselves most easily to a pre-post analysis, supplemented by qualitative accounts of policy decisions and practice changes. In other cases, such as mental health, one might expect a response over time (e.g., greater investment in community mental health programs, enhanced prevention programs).

4. Assembly of data on grants and other revenues made possible by the existence of Centre expertise and infrastructure.

5. Description of the extent to which deliverable reports are chronicled in the media. This is necessarily impressionistic and it is hard to assess how short-term media attention affects concrete responses to the findings.

A number of issues and controversies arise in applying the methods:

1. *Estimating the duration of impact*. Suppose that a deliverable results in a policy that reduces use of a particular service by $1 million a year. How long does the effect last, and how long should the savings be counted as ROI in the deliverable? As long as the policy is in place, presumably the savings will continue – the base could be permanently reduced. Conservatively, the estimates here restrict attribution to a ten-year period.

2. *Estimating avoided costs*. Similarly, accurately attributing avoided capital and/or operating costs to deliverable findings, and calculating the amount saved is fraught with uncertainty. It may be impossible to verify all of the factors contributing to a decision. Rarely will cause and effect be transparent. The rate of return depends on the method for calculating savings and the discount rate over time. For instance, if there is a decision not to build 30 hospital beds at a cost of $15 million, one could calculate the savings as:

   a. $15 million, one time
   b. The costs of borrowing $15 million at a certain interest rate – 5% would yield annual savings of $750,000 plus principal payments over the period of amortization
   c. One of a) or b), plus operating costs (e.g., 30 beds x 300 days per year x $400 in rural facilities = $3.6 million), less amounts spent in enhancing alternative services (assuming these are known)
   d. Zero, on the assumption that the money is not actually saved, but reallocated (to operations, or another department, or to pay down the provincial debt).

   Here we report a one-time capital cost saving only when there is consensus among decision-makers that a deliverable significantly influenced a decision not to proceed with a capital spending project that otherwise would have gone ahead. In such instances there is a strong likelihood that operating costs will be avoided, but it is also possible that these funds would have been spent elsewhere in the health care system. To avoid any risk of overstating savings,

we assumed that even where capital spending has been avoided, no operating savings accrue to the system as a whole.

3. *Older vs. more recent deliverables*. In theory, the older the deliverable, the more likely its full impact will have been experienced. At the same time, accurate recall of its impact and the environment in which it was delivered will diminish over time, and as always there are serious data limitations. Most of the people interviewed for this case study were not in their positions a decade ago. Conversely, more recent deliverables, especially if commissioned with an eye to the future, may not have had enough exposure and time to make their mark. The approach to these contingencies is as follows:

   a. For older deliverables (arbitrarily, those completed between 1997 and 2002) with the potential for major impact, we have sought financial, utilization, and/or health outcomes data, and evidence of policy outcomes, to offset any lack of recall by interviewees.
   b. For newer deliverables, we have assessed whether it is too early to expect a significant concrete impact, and have focused on whether the findings have affected awareness and policy discussions.
   c. We have selected a variety of older and newer deliverables to ensure that the report reflects both the range of commissioned work, and any trends in priorities.

4. *Policy vs. practice impact*. The endpoint of some deliverables is policy, while in other cases individual practice or organizational behaviour change may not entail formal policy-making. Where policy is the foreseeable outcome – and especially where there are binary (yes/no, stop/go) decisions, we have focused on decision-making. Where practice, behaviour, or change in approach is the expected result, we have sought utilization, cost, quality, and health outcomes data.

V. Information Sources

This report is based on the following information sources:
1. 15 interviews involving 20 people
2. MCHP deliverable reports from 1997 to 2007
3. Data on external research grants received by Centre researchers, from the University of Manitoba database
4. Data on health service utilization and costs related to Centre reports, from Manitoba Health and Healthy Living
5. Statements of responses to or the impact of MCHP deliverables and activities compiled by the Centre
6. Counts of citations of research published by core Centre investigators, from the ISI Web of Science database
7. Counts of newspaper articles referring to Centre reports, from the Canadian Newsstand database.

VI. Impact on Policy and the Culture of Decision-Making

Interviewees consistently noted that the Centre's contribution to the province has been far more than the sum of its parts. Two-thirds spontaneously referred to the impact of the Centre on the overall

culture: awareness of and sensitivity to research-based evidence, the importance and potential of data, and original insights and analyses that have changed the way people think.

On a 1 to 5 scale (with five the highest rating), interviewees indicated that the Centre has put new questions on the table and/or raised issues or generated discussion that would not have otherwise appeared, with an average score of 4.0. Just as importantly, respondents recognized that there are limits to the influence research is likely to have on significantly political processes.

> *It is useful to produce the evidence even if it doesn't influence decision-making as much as it should. It is very important to have evidence to confront the anecdotal and emotional. Must continue to educate the politicians and the general public on where resources should be allocated.[5]*

The interviewees were not systematically selected to represent the policy and health services delivery system. Bearing this in mind, the two deliverables mentioned most frequently as being influential were Estimating Personal Care Home Bed Requirements (Frohlich et al 2002) (9 mentions) and Patterns of Regional Mental Illness Disorder Diagnoses and Service Use in Manitoba: A Population-Based Study (Martens et al 2004) (6). These were followed by 4 mentions each of: Acuity of Patients Hospitalized for Medical Conditions at Winnipeg Acute Care Hospitals (Bruce 2001), Assessing the Performance of Rural and Northern Hospitals in Manitoba: A First Look (Stewart 2000), Projecting Hospital Bed Needs for 2020 (Stewart 2002), How Do Educational Outcomes Vary With Socioeconomic Status? Key Findings from the Manitoba Child Health Atlas 2004 (Brownell et al 2004), Assessing the Health of Children in Manitoba: A Population-Based Study (Brownell et al 2001), and A Needs-Based Funding Methodology for Regional Health Authorities: A Proposed Framework (Mustard et al 1998).

A major, separate source of influence is *The Need to Know* (NTK) team and interactions. Since 2001, this award-winning, CIHR-funded initiative has brought together the RHAs, Manitoba Health and Healthy Living and MCHP researchers to

> *The Centre is a very user-friendly, inclusive, credible organization.*

produce knowledge of key interest to planners, to build capacity among planners and researchers, and to work on models of dissemination and research application. Half the interviewees cited the existence and performance of the NTK team as a very positive Centre innovation that increases capacity, transmits knowledge, and supports the rural RHAs in particular in their efforts to compile and use evidence.

Interviewees rated the contribution of the Centre to the province's intellectual environment as the area of greatest impact (4.7 on a 5 point scale). They cited methodological innovations, award-winning knowledge translation activities, sustained interactions with boards, senior managers, and staff of the RHAs, and commitment to producing increasingly relevant and timely reports as the main elements of capacity building. Several RHAs reported that their evidence-oriented capacity has developed primarily as a result of the Centre's presence and outreach activities. All respondents viewed the Repository as essential and invaluable research and analysis infrastructure.

> *… [MCHP researchers]… have such a nice way of helping us understand research, and to apply the results in our decision-making… We are blessed to have such wonderful partnerships between our healthcare organizations and our research institutions in Manitoba.*

---

[5] Italicized quotes are from interviews, whose participants were guaranteed anonymity.

While there is no rigorous, systematic method for tracking the impact of Centre deliverables on policy development in Manitoba, over the years many decision-makers have communicated accounts of how the reports have been used. For example:

1. How Do Educational Outcomes Vary With Socioeconomic Status? Key Findings from the Manitoba Child Health Atlas (Brownell et al 2004) influenced the decision to launch the Community Schools Partnership Initiative by Manitoba Education, Citizenship & Youth in 2005-06.
2. Assessing the Health of Children in Manitoba (Brownell et al 2001) highlighted the extent of injuries and is cited as a continuing source of support for the prevention and health promotion efforts of Manitoba Health and Healthy Living.
3. The media release announcing the maximum allowable cost policy for drugs cited Pharmaceuticals: Therapeutic interchange and Pricing Policies (Morgan et al 2003) as a direct source of influence.
4. The Impact of Influenza-Like Illness on the Winnipeg Health Care System: Is an Early Warning System Possible? (Menec et al 2001) report was cited in the legislature as influential in efforts to reduce seasonal 'hallway medicine' in hospitals.
5. At least three RHAs have independently (i.e., apart from this exercise) reported their use of the reports on Projecting Hospital Bed Needs for 2020 (Stewart et al 2002), The Health and Health Care Use of Registered First Nations People Living in Manitoba: A Population-Based Study (Martens et al 2002), and Patterns of Regional Mental Illness Disorder Diagnoses and Service Use in Manitoba: A Population-Based Study (Martens et al 2004).

VII. Financial Impact on the Health Care System

Assessing the financial impact of the Centre proved to be the most difficult aspect of the case study, for reasons specified above. It is important to distinguish potential from actual impact. Centre deliverables have revealed a variety of ostensibly remediable inefficiencies in the system. Furthermore, even if some efficiencies are unachievable because of historical and political factors, the evidence may alter future decisions, reduce financial pressures in certain areas, and change the perspective of boards and communities. As one respondent put it, the expectation of the Winnipeg and particularly rural hospital bed use deliverables should not have been to reduce capacity, but to use it better.

The studies most frequently cited as having the most financial impact were the multiple reports of personal care home bed need and hospital bed needs projections. Interviewees working in specific program areas also mentioned deliverables germane to their portfolios. Several deliverables have studied the use of hospital beds and access to care. The overall findings have been that:

- Manitoba has historically had more hospital beds than most other Canadian jurisdictions.
- Despite Winnipeg bed closures, access has not been compromised, and more patients are being treated than ever before (Brownell et al 1999).
- Significant proportions of hospital days are used by patients who require an alternative level of care (ALC).
- Rural hospitals have particularly high proportions of ALC use.

The following table lists deliverables whose findings could in theory have the most significant potential impact on costs, with order of magnitude estimates. The column listing factors affecting implementation is at least as important as the dollar figures because it identifies some of the obvious challenges. Every health care system in the world operates at sub-optimal efficiency; evidence to that effect is a necessary but hardly sufficient condition for improvement. While deliverable findings and government policies and high-level decisions can be catalysts for progress, actions at the coalface ultimately tell the tale and organizational culture and behaviour are the keys to success. Effective change requires a combination of sound policy and incentives, organizational commitment and capacity, and in many cases, public support for potentially controversial initiatives.

To reiterate, these are rough estimates of potentially achievable savings or reallocations, not evaluated impact. The estimates are not based on unrealistic or lofty aspirations – e.g., 0% ALC rates in hospitals, or being the best performing jurisdiction in Canada. The estimate ranges are based on modest assumptions about proportions of inefficient practices that can be modified, and very conservative estimates of the savings achievable by substituting one form or venue of care for another. In prosperous countries, estimates of the proportion of ineffective, wasteful, and harmful health care spending range as high as 30% to 40%[6]. The cumulative maximum potential savings estimated here are about 3% of the Manitoba Health and Healthy Living budget.

---

[6] This is the estimate of renowned health quality champion Dr. Don Berwick; see "Improvement Tip: Find Muda [Japanese term for waste] and Root It Out," Institute for healthcare.

Table 1. Estimated Potential Financial Impact of Selected MCHP Deliverables

(See Appendix for derivation of estimated costs and savings)

DELIVERABLE	ASSUMPTIONS	CAPITAL COST AVOIDED		OPERATING COST AVOIDED		FACTORS AFFECTING IMPROVEMENT
		Est. Min.	Est. Max.	Est. Min.	Est. Max.	
High-Cost Users of Pharmaceuticals: Who Are They? (Kozyrskyj et al 2005)	Reduce no. of prescriptions by average of 1 or 2 among those with 10+ active prescriptions			$7,000,000	$14,000,000	Improved prescribing Regular reviews of high-use cases Practice-level innovations and quality improvement initiatives
Patterns of Health Care Use and Cost at the End of Life (Menec et al 2004)	Reduce LTC residents' hosp days from 16,000 to 14,000 or 12,000 Reduce home care clients' hosp days from 79,000 to 71,000 or 63,000			$2,000,000	$4,000,000	Expanded palliative home care LTC policies and capacities to reduce transfers Public discussion of and uptake of advanced directives Greater intersectoral coordination of responsibilities
Pharmaceuticals: Therapeutic Interchange and Pricing Policies (Morgan et al 2003)	Substitute lower cost ACEI of equivalent therapeutic value Use ACEI before A2RA			$5,000,000	$7,250,000	There are a variety of approaches to managing drug utilization and costs. Other drug classes may be prioritized. Ability to counter predictable opposition from parts of Rx industry
Estimating Personal Care Home Bed Requirements (Frohlich et al 2002)	Mid-range projections are best achievable scenario Due to ongoing shift away from residential care, etc. only 300 of 1200 new beds under 'recent use pattern' scenario would have been built in absence of deliverable	$0	$55,500,000	$16,425,000	$31,025,000	Enhanced home care, supportive housing Policies and funding formulas that encourage community care Public support for shift to community care options
Projecting Hospital Bed Needs for 2020 (Stewart et al 2002)	Attribute 10% to 20% of the difference in hospital days between the current use projections and the trend-line projections to the deliverable			$10,000,000	$20,000,000	Policies and practices designed to reduce the ALC rate in hospitals Continued transfer of procedures from inpatient to outpatient Efforts to refine and implement appropriateness criteria Greater intersectoral coordination facilitated by funding practices and incentives
Acuity of Patients Hospitalized for Medical Conditions at Winnipeg Acute Care Hospitals (Bruce et al 2001)	69,000 ALC days Winnipeg 2006 20% to 40% reductions feasible via substitutions			$3,450,000	$6,900,000	Enhanced home care for patients awaiting LTC panelling or placement Flexible or short-stay LTC beds to accommodate surges in long stay hospital patients Policies and practices designed to reduce the ALC rate in hospitals Greater intersectoral coordination facilitated by funding practices and incentives Standardized discharge protocols that reduce variations in practice
Assessing the Performance of Rural and Northern Hospitals in Manitoba: A First Look (Stewart et al 2000)	Based on utilization and patterns in SK, ALC rates of at least 30% 25% to 50% reduction in ALC rate feasible via substitutions $301 million hospital budget 2006-07 for all regions excluding WRHA and Brandon 5% to 10% savings			$15,000,000	$30,000,000	
TOTALS		$0	$55,500,000	$58,875,000	$113,175,000	

In contrast to these conservative estimates, a Treasury Board submission estimated preventable capital costs of $400 million in the personal care (nursing) home sector would be possible by pursuing an 'aging in place' program that shifted the locus of care to supportive housing. Reducing the proportion of ALC days in rural hospitals to 20% would result in tens of thousands of avoided hospital days and millions of dollars annually.

Some real and major changes in resource allocation and philosophy of care are impossible to quantify because the effect of the deliverables has been more broadly cultural than specific to any program or service. Many interviewees said that the nursing home bed needs and rural hospital use research changed expectations and priorities in at least some RHAs. The focus has shifted from institutions to programs such as mental health and primary health care – both of which were the subject of major deliverables. There is less pressure on the government to build new facilities.

A number of respondents highlighted the Martens et al (2004) report on mental health prevalence and service use as an influential deliverable. RHA mental health spending has risen at uneven but substantial rates (an average of nearly 12% compounded annually over the past five years), but it does not appear that the trend changed after the release of the report[7]. Greater awareness of mental health issues may influence practices that do not show up as changes in resource allocation.

The Morgan et al (2003) report on drug utilization is one of the few where it is possible to compare utilization and costs prior to and after its publication. There appears not to have been any change in trends in ACEI and A2RA utilization associated with the release of the report, although without a control group it may be premature to assume no impact. ACEI use has levelled off following a period of explosive growth. By contrast, the

From Insight to Impact: A Case Study

In the early 1990s there was a plan to replace 2 aging hospitals in Morden and Winkler with a single facility serving both communities and surrounding area. The initial plan was for more beds than were eventually built. During the capital planning process, Manitoba Health officials began to examine rural hospital bed needs more closely given Manitoba's historically abundant supply. At the same time, the Centre produced its first report on how rural hospitals were being used – or not. The Centre presented its findings to the largely rural government caucus of the day.

Centre staff then worked closely with Manitoba Health officials to refine the assessment of needs and assess how the study findings, along with other evidence, could be applied to the planning process. The internal government analysis produced a recommendation to scale down the number of beds.

Ultimately the decision was to scale down the new hospital to 94 beds (25 rehabilitation), and it opened in May 2001 at a cost of $37.6 million, a decrease of at least $5 to $8 million. Assume a cost of operating at least 15 beds not built of $300/day and an occupancy rate of 80%, of which $175/day is a net saving. Over the past seven years the cumulative operating cost saving has been about $5 million.

It is unknowable whether the Centre deliverable is entirely or partly responsible for the savings. Undoubtedly the evidence strengthened the case for reducing the bed complement, and the presentation of the broader study to the government caucus helped facilitate a difficult political decision. It does seem reasonable to claim that the deliverable provided necessary evidence for revisiting the facility plans. It took a receptive and courageous political environment to create the sufficient conditions for making a "rational" decision.

---

[7] There is little consistency among regions and there are anomalies in the data, e.g., Brandon RHA reports spending more than Winnipeg RHA on mental health, attributable to the historical presence of a large inpatient psychiatric facility which closed in 1998.

A2RA trend is sharply upward before and after the study: a 32% compounded annual growth rate from 1998-99 through 2005-06. The substantial cost savings resulted from generic substitution enabled by the Manitoba Interchangeable Formulary.

Treasury Board makes the major decisions on total health care funding. MCHP deliverables are frequently cited in Treasury Board submissions. Rural RHAs in particular report that both the deliverable findings, and the capacity to acquire and use evidence to inform decisions and submissions influence their priority-setting and budget processes. This is highly relevant, given that the provincial government mandates that each RHA submit evidence-based five-year strategic plans to Manitoba Health and Healthy Living. Thus MCHP and the NTK team structure have been key elements in RHAs to fulfill this mandate.

Finally, respondents reported that the indicators atlas reports (RHA Indicators Atlases in 1999, 2003 and 2008; First Nations report in 2002; Mental Health Atlas in 2004, Sex Differences in Health Atlas in 2005; What Works report in 2008; Child Health Atlases in 2001, 2004, 2008) have strongly influenced assessments of how well programs and services served the community. Numerous deliverables have documented mismatches between patterns of care and utilization and health care needs. The atlases vividly show variations in population and geographic needs, which has led to discussions about how services might be reconfigured. These deliberations are important harbingers of change even though their origins and impact may be somewhat invisible.

VIII. External Revenue and Research Efficiency

The Centre is not just a provider of services; it is a university-based institution with a measurable impact on the size and reputation of the Manitoba research community. Centre researchers, supported by the unique infrastructure developed over the years, compete successfully for grants from external granting agencies, as do non-Centre researchers who use Centre databases and expertise. These revenue categories are[8]:

2000 – 2006 MCHP researchers
$11.3 million
2000 – 2006 non-MCHP researchers
$13.2 million
1991-1999 MCHP researchers and collaborators
$21.4 million
TOTAL
$45.9 million

From 1991-92 through 2005-06 the Centre received approximately $26.2 million from Manitoba Health and Healthy Living. No more than a small percentage of the externally funded research would have been feasible in the absence of the Centre given the centrality of the Repository to the research enterprise. One could then infer that for each dollar Manitoba Health and Healthy Living has invested in the Centre, approximately $1.70 has come to the province in external research support. Put another way, the province has realized a net gain of about $20 million on this measure alone. Research grants are extremely efficient job creators – on the order of $50,000 to $75,000 per person-year of employment.

---

[8] Prior to 2000 it is not possible to identify which grants were exclusively or partly received by Centre researchers, hence the distinction in the figures presented.

Over the 15 year period, assuming that at least $15 million in external funding paid research and technical salaries, at least 200 person-years of high-quality work are directly attributable to the external revenue-generating capacity of the Centre.

Fig. 1 External Grants vs. MB Health Revenues, MCHP, 1991-2006

$ Millions
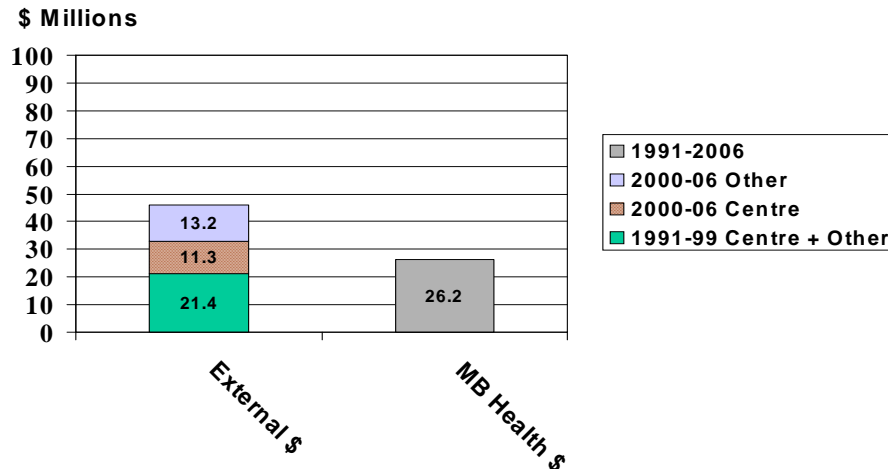
The economic impact of externally funded research does not stop at the grants alone. Health care, social science and business students have all been enriched by Centre expertise and resources. There are broader, if indeterminate economic spinoffs from the presence of scientific and technical expertise. Because the Centre has pursued a knowledge translation and capacity enhancement agenda, there is a growing community of research producers and consumers throughout the health care system – particularly outside Winnipeg, where interviewees stated that such developments would have been impossible without the Centre. Similarly, several government respondents indicated that their own knowledge, expertise, and analytic capacity have been strengthened by the Centre, and welcomed the prospect of more interaction on a number of fronts (NTK team, joint seminars, participation on working groups for deliverables, etc.).

The extent to which the Repository enhances both the quality and efficiency of the research enterprise is surmised below:

1. It expands research possibilities, especially studies linking the determinants of health to the use of health services and health status outcomes. The school completion project is an example in Manitoba, but Canada has become a world centre of excellence in population health research because of the growing capacity to link databases across sectors.
2. It makes the research enterprise more efficient by building in quality control processes that reduce or eliminate the need for intensive effort in assembling and cleaning data for each study.
3. It supports the creation of tools such as templates for mining and reporting on data that can be used by a wide variety of analysts and researchers, reducing the effort required to extract usable output.
4. As data holdings expand, the potential for interdisciplinary research increases. As interviewees observed, the Centre holdings are now of growing interest both within (clinician researchers)

and outside the health sector, and research and analysis can now support the search for greater understanding of the factors affecting health and potential solutions to complex problems.

IX. Reputation and External Impact

The Centre is considered to be more than the sum of its deliverables, and the credibility and influence of its work are not limited to Manitoba. Reputation is an important, if difficult to quantify, aspect of the Centre's performance and influence. There are several aspects to and implications of its reputation, as follows:

1. ***Scientific and methodological reputation within Manitoba.*** Interview respondents were near unanimous in their praise of the Centre's intellectual capacity, methodological sophistication, and ability to produce comprehensive and careful reports. Even the two respondents who took issue with some of the technical qualities of certain deliverables confirmed that the Centre is deservedly held in high regard.

2. ***Integrity and independence.*** It was considered a great strength of the Centre that it is, and is perceived to be arm's length from government, strong academically, and committed to discovering truth rather than serving a particular policy or political agenda. Several respondents indicated that it would be impossible to build similar capacity within government because it is difficult to attract, retain, and support researchers in that environment, and also observed that in-house products would invariably be considered biased in some quarters.

3. ***Reputation as a partner, capacity builder, and communicator.*** Again, the Centre is highly regarded as a partner to the health care system and the government. Several respondents noted that in recent years the Centre has become even more responsive to the needs of the system and has enhanced its efforts to build capacity in the RHAs. Several explicitly cited the opportunities to influence the Centre's agenda and participate on working groups as important

> *We have 2 people who 3 times a year spend 2 days at the Centre. They have grown in their understanding, come back highly motivated to share information and methods, a camaraderie has developed under [the Director's] nurturing and team-building.*

contributions. All who commented stated that the Centre's reports were highly readable and accessible to a wide audience (particularly the Summary Reports). The NTK team has built a great deal of capacity and support for research and analysis.

4. ***Reputation and presence in the media.*** This is more difficult to quantify, but there is evidence of a substantial and ongoing presence. A search on the multiple ProQuest databases with the keywords 'Manitoba Centre for Health*' or the names of Centre researchers most heavily involved in producing

> *The Centre is extremely highly regarded in other provinces. It brings a level of prominence to Manitoba in research.*

deliverables over the past decade or so yields about 500 citations in Canadian newspapers and other periodicals since 1993.

5. ***Reputation in the scientific and academic community.*** The credibility of the Centre, both within and outside Manitoba, ultimately depends on the excellence of its work. A widely used measure of scientific and methodological impact is the number of citations to a body of work in the scientific literature. A search of the ISI/Web of Science database yielded at least 6,500 discrete[9] citations since 1975. Individual articles on the use of administrative bases and other

> *Overall, you [the Centre] are obviously the best we have [internationally]. I did enjoy the Medical Care issue (Academics at the Policy Interface, 1999; 37). In fact, I've cited it as one of the motivations for our recent proposal for funds for a five-year thematic research plan [in Western Australia].*

pioneering methodological work have been cited up to 400 times – a remarkably high count for social science literature. Centre researchers have made countless presentations within and outside Manitoba at conferences and seminars. Several have won salary awards from prestigious national granting agencies (NHRDP, CIHR) and the Manitoba Health and Healthy Living Research Foundation. A number of respondents said that the existence of the Centre is a magnet for new faculty at the University of Manitoba.

6. ***Impact on policy and practice in other jurisdictions.*** The Centre has compiled accounts and communications of where its work has been reported to have influenced developments in other jurisdictions. Among these are:

   a. The BC and Ontario governments have cited the influenza report (Menec et al 2001) as an influence on their decisions to expand immunization programs.

   b. The Hawaii Health Information Corporation's 2004 report 'Maui Bed Needs Study, 2005-2025' stated that its model relied heavily on the Centre's Projecting Hospital Bed Needs for 2020 report (Stewart et al 2002).

   c. The report on physician supply and use (Watson et al 2003) has influenced the work of the Canadian Medical Association, the College of Family Physicians, and the federal government in health human resource planning.

X. Discussion and Conclusions

Notwithstanding the caveats and limitations inherent in estimating the return on the investment in the MCHP, its impacts have been demonstrated in numerous ways, including reputation, research revenues and productivity, varying influence on policy and system management, and a major cultural and intellectual influence on the Manitoba environment. The quantifiable ROI from research grants alone is close to 200%, but the real *impact* is likely to be far greater.

The case study allowed us to test the limits of existing methodologies in estimating the impact of a considerable variety of HSR studies. The estimates are deliberately conservative and we do not claim that they are precise. Attribution is difficult even where research findings and subsequent decisions appear to align.

Perhaps the most important conclusion is that it would be an oversimplification to characterize HSR as a purely applied, utilitarian enterprise with predictable impact. Rather, it is a form of knowledge

---

[9] A discrete count means that an article co-authored by more than one Centre researcher would be counted only once. The search was limited to the non-duplicating citations to the work of NP Roos, LL Roos, P Martens, M Brownell, C Black, V Menec, and CA Mustard (for Black and Mustard limited to the years when active with the Centre or where the article was obviously related to a Centre product).

generation whose findings are relevant to a diverse and complex sector, fraught with interests, traditions, uncertainty, technical limitations, and political dynamics. Health and health care are prone to information overload; misalignment between needs, resource allocations, and the capacity to intervene successfully; and values conflicts. Rare is the dilemma that waits for and relies entirely on an HSR solution.

Nonetheless, with effort it is possible to identify where at least some types of HSR are and are not influential, and there is reason for optimism that methods can be refined if impact evaluation is built into the conception and conduct of the research and its fate following completion. Mixed methods can reinforce and increase confidence in estimates of impact. Acknowledgement that not all impact can be quantified, and that intellectual and cultural influence can be far more powerful than immediate resource impact, should inform further methodological initiatives. Financial impact will always be debatable but it should be possible to develop more consistent approaches to estimates, and as long as the methods and assumptions are transparently disclosed, the cause is advanced.

XI. Recommendations

1. Funders, producers, and consumers of HSR should recognize that it spans the range from fundamental to applied science, and its potential to create new knowledge and raise new issues is as important as its potential to influence short-term decision-making.

2. HSR projects should be categorized by their nature and potential impact, to clarify their intent and potential domains of influence, refine their methodologies, identify key stakeholders and target audiences before the work is undertaken.

3. For commissioned HSR projects in particular, the preparatory stages should identify the potential short and longer term impact of hypothetical findings, a plan for disseminating results, and responsibility for translating them into policy and resource allocation options. Building these activities into project scoping and design would highlight and maintain a focus on potential impact and create realistic expectations for uptake and application.

4. Where feasible and appropriate, HSR projects should build in plans for evaluation of impact and should identify the data (qualitative and quantitative) required to produce reliable estimates on various dimensions.

5. Researchers, policy-makers, and administrators should continue to refine concepts and methods for quantifying the financial impact of HSR findings, and particularly:
   a. Attribution
   b. Duration
   c. Discount rates
   d. Estimates of cost avoidance.

6. Health services researchers and organizations, granting agencies, research funders, auditors, program evaluators, and others should create forums and structures to work systematically to refine impact and ROI methods, compile case studies, and clarify accountabilities.

References

Agency for Healthcare Research and Quality, 2002. [Online] Available at: http://www.academyhealth.org/hsrproj/definitionofhsr.htm

Balas E.A., Boren, S.A., 2000. *Managing clinical knowledge for healthcare improvements*. Germany: Schattauer Publishing, p. 65 – 70.

Brownell, M., et al, 1999. Monitoring the Winnipeg hospital system: 1990 – 91 through 1996 – 97. [Online] Available at: http://mchp-appserv.cpe.umanitoba.ca/deliverablesList.html

Brownell, M. et al, 2001. Assessing the health of children in Manitoba: a population-based study. [Online] Available at: http://mchp-appserv.cpe.umanitoba.ca/deliverablesList.html

Brownell, M., et al, 2004. How do educational outcomes vary with socioeconomic status? Key findings from the Manitoba Child Health Atlas. [Online] Available at: http://mchp-appserv.cpe.umanitoba.ca/deliverablesList.html

Bruce, S., et al, 2001. Acuity of patients hospitalized for medical conditions at Winnipeg acute care hospitals. [Online] Available at: http://mchp-appserv.cpe.umanitoba.ca/deliverablesList.html

Buxton, M., Hanney, S., & Jones, T., 2004. Estimating the economic value to societies of the impact of health research: a critical review. *Bulletin of the World Health Organization,* 82, p. 733 – 739.

Cabana, M.D., et al, 1999. Why don't physicians follow clinical practice guidelines? A framework for improvement. *The Journal of the American Medical Association*, 282 (15), p. 1458 – 1465.

Frohlich, N., et al, 2002. Estimating personal care home bed requirements. [Online] Available at: http://mchp-appserv.cpe.umanitoba.ca/deliverablesList.html

Hanney, S.R., et al, 2003. The utilisation of health research in policy-making: concepts, examples and methods of assessment. *Health Research Policy and Systems*, 1 (2), p. 1 – 28.

Hovland, I., 2007. Making a difference: M & E of policy research. London: Overseas Development Institute.

Innvaer, S., et al, 2007. Health policy-makers' perceptions of their use of evidence: a systematic review. *Journal of Health Services and Policy Research*, 4, p. 239 – 244.

Institute of Medicine (IOM), 2001. *Crossing the quality chasm: a new health system for the 21st century.* Washington, DC: National Academy Press.

Kozyrskyj A, Lix L, Dahl M, Soodeen R. High-cost users of pharmaceuticals: Who are they? [Online] Available at: http://mchp-appserv.cpe.umanitoba.ca/deliverablesList.html

Kuruvilla, S., et al, 2006. Describing the impact of health research: a research impact framework. *BMC Health Services Research*, 6, p. 134 – 151.

Martens, P., et al, 2002. The health and health care use of registered First Nations People living in Manitoba : a population-based study. [Online] Available at : http://mchp-appserv.cpe.umanitoba.ca/deliverablesList.html

Martens, P., et al, 2004. Patterns of regional mental illness disorder diagnoses and service use in Manitoba: a population-based study. [Online] Available at: http://mchp-appserv.cpe.umanitoba.ca/deliverablesList.html

Martens, P., et al, 2008. What works ? A first look at evaluating Manitoba's regional health programs and policies at the population level. [Online] Available at: http://mchpappserv.cpe.umanitoba.ca/deliverablesList.html

Menec, V., et al, 2004. Patterns of health care use and cost at the end of life. [Online] Available at: http://mchp-appserv.cpe.umanitoba.ca/deliverablesList.html

Morgan SG, Kozyrskyj A, Metge C, Roos NP. Pharmaceuticals: Therapeutic interchange and pricing policies. [Online] Available at: http://mchpappserv.cpe.umanitoba.ca/deliverablesList.html

Mustard, C., et al, 1997. A needs-based funding methodology for regional health authorities : a proposed framework. [Online] Available at : http://mchp-appserv.cpe.umanitoba.ca/deliverablesList.html

Soper, B., Hanney, S.R., 2007. Lessons from the evaluation of the UK's NHS R&D Implementation Methods Programme. *Implementation Science*, 2, p. 7 – 19.

Stewart, D., et al, 2000. Assessing the performance of rural and northern hospitals in Manitoba: a first look. [Online] Available at: http://mchp-appserv.cpe.umanitoba.ca/deliverablesList.html

Stewart, D., et al, 2002. Projecting hospital bed needs for 2020. [Online] Available at : http://mchp-appserv.cpe.umanitoba.ca/deliverablesList.html

Paper Appendix: Methods for Deriving Cost and Savings Estimates

Patterns of Health Care Use and Cost at the End of Life (Menec 2004)
LTC bed cost/day: $125
Hospital bed cost/day for LTC transfers: $325
Home care cost/day (5 x avg. in report): $125
Hospital bed cost/day for home care transfers: $325

These figures are conservative estimates of cost savings achievable by substituting LTC and home care for hospital care in that the hospital per diem costs are likely underestimated.

All decedents in Manitoba account for 21% of total health care costs in their last year of life – on the order of $800,000,000 per year. The estimated maximum achievable saving of $4,000,000 is only 0.5% of the total – again, extremely conservative.

Pharmaceuticals: Therapeutic Interchange and Pricing Policies (Morgan 2003)
Taken directly from the report – no separate assumptions made.

Estimating Personal Care Home Bed Requirements (Frohlich et al 2002)
Take the mid-range estimate from the report on expected bed needs for 2020. Potential to take 500 beds out of circulation. Recent use model projected 1200 new beds needed by 2020. Assume up to 25% of these beds would have been built if deliverable had not studied issue and reported high use in Manitoba relative to other Canadian and international jurisdictions.

Minimum capital cost avoided: assume 0 new beds would have been built.
Maximum capital cost avoided: assume 25% of 1200 (300) x $185,000/bed.
Operating cost savings from substituting home care for LTC:

Per diem LTC costs:	$125
Per diem home care costs (3 x provincial average):	75

Minimum operating cost savings: 900 LTC residents looked after in community (1200 vs. 300 new beds).
Maximum operating cost savings: 1700 LTC residents looked after in community (1200 vs. –500 beds).

Projecting Hospital Bed Needs for 2020 (Stewart et al 2002)
Begin with 500,000 patient-days/year difference between 1997-98 utilization patterns and projected to 2020, and trend line projections. Attribute 10% to 20% (50,000 to 100,000 days) of the difference to the impact of the various MCHP studies on hospital use and bed needs projections. Assume net saving of $200/day from substituting other forms of care for hospital care (very conservative). Assume no new beds would have been built regardless (very conservative).

Acuity of Patients Hospitalized for Medical Conditions at Winnipeg Acute Care Hospitals (Bruce et al 2001)

In 2006 there were 123,000 ALC days in Winnipeg hospitals. Excluding rehabilitation days left 69,000. Based on a combination of Manitoba data and detailed cost estimates of hospital days from the Calgary Health Region, it is estimated that:

> ➢ a medical bed costs about $800/day
> ➢ an ALC day costs $400/day
> ➢ all other forms of care cost $150/day on average (almost certainly too high in that many patients do not require significant care on discharge, PCH beds cost less than $150/day, etc.)

Therefore the per diem savings are $250/day. It is assumed that no more than 20% to 40% of the ALC days can be eliminated – a very conservative assumption in that a number of metropolitan areas have much lower bed ratios than Winnipeg.

Pillar IV: Population and Public Health Research

Assessing the Return on Canada's Public Investment in Population and Public Health Research: Methods and Metrics

Alan Shiell PhD

Professor, AHFMR Health Scientist and

CIHR/PHAC Chair in Public Health Economics

Population Health Intervention Research Centre

University of Calgary

Erica Di Ruggiero MHSc RD

 Associate Director

CIHR Institute of Population and Public Health

Department of Public Health Sciences

University of Toronto

Declaration of Interest

Both authors benefit from the support of the CIHR Institute of Population and Public Health. Alan Shiell has received operating grants from the CIHR, holds a CIHR/ PHAC Chair in the Economics of Public Health and is a member of a research centre established with CIHR funds from the Centres for Research Development program. Erica Di Ruggiero is the first Associate Director of the Institute of Population and Public Health.

Summary

Since its inception in 2000, the Canadian Institutes of Health Research (CIHR) has spent more than $6 billion of public funds on all forms of health research. The research this money supports covers a wide spectrum of activity ranging from bench-based science in the laboratory to population-based studies aimed at better understanding of the complex causes of ill health and the merits of different ways of preventing disease and promoting good health. This is money that could be used in other ways, and could perhaps improve health directly if spent on health care. It is instead spent on research in anticipation that the benefits it will bring in the future more than offsets the costs we all pay today. Recognizing the need for transparency and accountability for the use of this money, research funding bodies including CIHR, have looked for ways to evaluate the impact of research. Here we focus on what CIHR can do to evaluate the impact of its funding of population and public health research.

Several frameworks for evaluating the impact of research have been developed and applied around the world. We review two of these: the 'return on investment approach', which purportedly provides an indication of how the monetary value of the health gains that follow from successful implementation of research findings compares with its costs, and the 'payback approach', which provides a more comprehensive description of the impacts of research funding beyond those that can be valued in monetary terms. We also discuss some of the challenges that will be encountered. This includes the need to better define the scope of population and public health research so that the total investment can be quantified. The biggest challenge however will lie in correctly attributing to CIHR's investment in population and public health research an appropriate proportion of any impact that has been realized.

Given the challenges we face, we find that each of the two approaches to evaluating the impact of research has some merit and both should be applied, albeit with caution. The return on investment approach imposes few additional information requirements, though some analysis is needed. The insights it offers come at low cost therefore, and while the results are unlikely to carry universal agreement, they might provide a crude indication of the net-value of research - its sign if not its magnitude. This may have some merit politically. More importantly, the exercise will prompt much needed discussion on the attribution problem mentioned above and it will have served its purpose if it achieves nothing more than agreement that the approach is too crude to generate useful results. The payback approach is better suited to the Canadian context, not least because it is more easily adapted to suit the multiple objectives and the stage of development of population and public health research funding in this country. We discuss several metrics that can be used to operationalize the payback approach including both traditional measures such as citation rates, as well as new measures that are needed to capture the novel and sometimes complex objectives of CIHR's investment in population and public health research.

The payback approach is more time-intensive than the return on investment approach and is therefore likely to be more costly. Costs can be reduced however whilst maintaining the advantages offered by the payback approach, by combining broad-brushed, more frequent evaluations using just the readily available indicators, with occasional, more intensive evaluations requiring new data or additional analyzes. Combining metrics and methods in this way allows one to triangulate the results and thereby examine the validity of conventional measures of research output. The payback approach does not provide a perfect solution to the attribution problem, but the framework it provides allows one to examine *whether* the impact of particular research activities depends on previous research, if not the actual extent to which this happens.

In conclusion, any evaluative exercise such as that proposed here can be challenging because of the danger that we misrepresent and understate the value of the activities being appraised. All health research is vulnerable to this problem because of the difficulty attributing valued outcomes to particular research activities. Population and public health researchers need not fear the exercise though. We have seen how valuable population and public health research can be in the reduction in deaths and disability from Sudden Infant Death Syndrome, from infectious disease, and from tobacco-related diseases to name just three examples. Canadians believe in the merits of health research and happily support it. They deserve to know that it is having the expected benefit.

1.   Introduction

Twenty years ago more than one in every thousand babies born alive in Canada died of Sudden Infant Death Syndrome (SIDS). Today, the rate is less than one-half of this. Much of the reduction in SIDS happened very quickly during the 1990s following the widespread introduction of an intervention to reduce the risk of SIDS. The intervention was a simple one: parents and child-care providers were encouraged to place infants on their back to sleep. The research that led to this breakthrough did not take place in the laboratory or the clinic, but was 'population health' research. It involved painstaking examination of the interactions among risk factors and differences in the rates of SIDS from population-level studies in several countries to identify putative pathways and possible entry points for intervention, supported by further population-level studies to evaluate the impact of new preventive policies, including the 'Back to Sleep' campaigns that were implemented around the world, based on that evidence (see case study, p. 21).

The evidence that fuelled the reduction in SIDS came from several different countries: from the USA, Australia and New Zealand, the UK, the Netherlands and the Nordic countries, but not Canada. One possible reason for this is the relative lack of spending on population health research in Canada at the time. This has now changed. Since its inception in 2000, the Canadian Institutes of Health Research (CIHR) has spent more than $6 billion of taxpayers' money on research aimed ultimately at improving the future health and well-being of Canadians and others. This is not much in comparison with the $150 billion we spend *each year* on health care, but the amount of public money spent on all forms of health research increased in real terms throughout the first half this decade (Leaders Forum Steering Committee 2004). Of this $6 billion, perhaps as much as $450 million has supported the type of population-level research that led to the substantial reduction in SIDS around the world.

As the SIDS example illustrates, the benefits of research spending can be substantial. Research is a risky activity, however, and not all research is successful. Furthermore, the money allocated to research could otherwise go directly to support health care. It is instead invested in research in anticipation that the benefits this will bring to us all tomorrow more than offsets the costs we pay today. Recognizing the need for accountability and transparency for the use of these public funds, health research funding bodies in Canada and overseas have begun to examine how they might best assess the impact or value of the research activities that they support: in short to examine whether our expectations of benefits exceeding costs have foundation. A number of methods have been developed for assessing different components of the impact of health research. The results to date tend to suggest that the value of health research does exceed its costs. We review two of these frameworks to assess their suitability for assessing the value of the investment that CIHR has made since 2000 specifically in population and public health research.

We begin by describing what is meant by 'return on investment' and by discussing the merits of the two approaches to evaluating the impact of health research (Section 2). We then discuss the scope of population and public health (PPH) research in Canada, how much has been spent by CIHR in this research domain and with what intent, and the fit between CIHR's objectives with respect to its funding of PPH research and the two methods for assessing the impact of health research reviewed previously (Section 3). In Section 4, we consider how one might proceed to evaluate the impact of CIHR's investment in PPH research, the measures available and the anticipated challenges. In Section 5 we draw conclusions.

2. Frameworks for Quantifying Return on Investment

2.1 What is Meant by Return on Investment?

'Return on investment' typically refers to the net income (or profit or surplus) generated from investment in an asset of some sort, expressed as a percentage of the cost of that investment. The term is most commonly conceptualized in financial terms, where the rate of return is equivalent to the rate of interest earned on the investment. One reason for this association between return on investment and finance is that in the business sector, where return on investment is most frequently calculated, profit is the most important (if not the only) benefit. Furthermore, other outcomes, such as customer loyalty or brand recognition for example (if they are at all relevant to the calculation at hand) can be relatively easily translated into monetary terms.

The notion of return on investment need not be restricted to considering only financial benefits, however. Conceptually, we can think of 'return on investment' as referring to the net value of whatever we gain in the future as a result of expending resources (time and effort) today: and this concept applies no matter what it is that generates value providing that it can be expressed in dollar terms. In this more general sense, precisely what counts as a valued output can be adapted to fit contexts other than the business sector. The return on investment can therefore include social as well as private benefits, and outputs that are more difficult to value in monetary terms. For example, we can use economic methods to place a monetary value on health (which we discuss later) and so even the value of improvements in health that flow from the effective implementation of research findings can be included in an estimate of return on investment.

By interpreting the concept of return on investment in this way we gain the power that comes from the use of economic rhetoric. For example, the notion of 'investing' in research conjures up a more favourable, prudent image than 'spending', which is often deemed extravagant, even though the only real difference is the time frame over which the benefits of the expenditure is enjoyed. With spending the benefits are enjoyed today, whereas with investment the benefits are deferred until sometime in the future. The advantage comes at a price however, since we take on the 'baggage' associated with economic rhetoric and economic methods, and this may occasionally outweigh the benefits. In reflecting on possible approaches to valuing CIHR's investment in health research, the participants in an expert meeting held in 2005 accepted the more inclusive notion of return on investment discussed here. They also acknowledged however, that economic language is so closely aligned in many people's eyes with a narrow range of financial benefits that use of economic terminology is most likely counter-productive (Canadian Institutes of Health Research 2005). Meeting participants preferred instead to talk of the impact of research or its social value when referring to benefits that extend beyond those easily valued in monetary terms.

In what follows we employ the term 'return on investment' either where we refer to the work of others who use this term, or where it is possible to express the important outcomes of research in monetary terms. We refer to the impact of health research where we want to be inclusive of a broader range of outputs, not all of which are readily expressed in monetary terms. This includes intermediate outcomes such as research-capacity building, researcher-decision maker linkages, and production of evidence-based guidelines: that is outputs that are not valued in their own right but that facilitate the production of valued outputs. Finally we refer to 'social value' to indicate a judgement about the net worth of these and other research outputs, which may be qualitative rather than quantitative: that is whether the value of all that is gained through research exceeds its costs.

When evaluating the impact of research, one is interested in how the value of whatever is gained through research activity compares with the cost of that research. If research is effective, then it generates new knowledge that helps us better understand a phenomenon or enables us to act and respond to a problem more effectively. We are concerned not just in the production of new knowledge, but in its value to someone other than the researcher. The value of new knowledge can be *instrumental*: it has utility to someone, or it can be *intrinsic*: where the acquisition of knowledge is valued for its own sake.

Whatever the nature of the value generated by research, it is important to distinguish measures of research *output* (what has been produced by the research endeavour) and evidence of research *outcome* or the value-added. The two are obviously related since there can be no value-added without prior research output, but output is only a necessary condition for generating value, it is not a sufficient condition. Research might generate new knowledge but offer little return on investment. The new knowledge may relate to an issue of little or no importance or might not make a difference to whatever course of action one subsequently takes. New knowledge that is not translated into health-improving changes in policies, programmes or products can have intrinsic value but, other things being equal, generates less return than would otherwise be the case. We need to also distinguish – at least conceptually – the value of population and public health *research* from the value of population and public health *practice and policy,* which may be informed by or influenced by research but not always. The latter may be highly effective and highly valuable in promoting health and in contributing to the economic and social well-being of Canadians, but this does not mean that all PPH research is necessarily as valuable. We happen to believe that most PPH research is valuable, but the challenge we face is to demonstrate that.

There are two other key considerations that must be taken into account when we think about the social value of health research. If spending on research is to generate new and value-added knowledge, then there needs to be pre-existing *research capacity* ready to receive and use the research funds: that is there needs to be a critical mass of researchers with the requisite human capital needed to design, execute, analyze, interpret and disseminate the results of the funded research. Second, if the results are to be translated effectively into policy and practice then there also needs to be pre-existing *receptor capacity*: that is a research-literate policy and practice community who is engaged early and often throughout the research process and is receptive to its results. Thus, before one can reap the benefits of research, it may be necessary to invest first in building research and receptor capacity.

Finally, assessments of the social value of health research are often retrospective exercises. As such, one is trying to compare actualized value with research costs incurred in the past. Such information is useful

in assessing whether past investment decisions were sensible ones or not: that is in addressing issues of accountability, but one cannot reliably infer from such an exercise the value of current or future investment decisions. A slightly different approach is needed to prioritize research funding decisions: one that looks forward to estimate the expected value of the information likely to be generated by the particular research proposal being evaluated (Ginnelly et al 2005).

2.2   Past Attempts to Assess the Social Value of Investment in Health Research

There have been several attempts to quantify different aspects of the impact of society's investment in health research. In a review of some of this literature – that focusing on economic returns only - Buxton and colleagues (2004) discuss the results of 23 studies conducted in seven countries (predominantly the USA) dating back to 1967. The studies varied in their scope (national versus international), in their focus (ranging from research into specific diseases or treatments up to national research programmes) and in the methods they each adopted. All however, limited their evaluation to benefits easily measured in monetary terms.

Four different approaches to conceptualizing and measuring the economic impact of health research were identified: each somewhat partial in its approach. The first group of studies focused on the impact that new technologies have on the costs of health care. The second group of studies examined the effects that a healthier workforce has on economic production. The third group were studies quantifying the income and employment generated from new product development and commercialization, especially from investment in the pharmaceutical industry. The fourth group considered the intrinsic value to society of improvements in health brought about by the implementation of research findings.

The authors of the review found no consistent methodology employed across the studies reviewed, and no study was comprehensive in its coverage. They concluded that the final approach held out the most promise to those concerned with assessing the economic value of health research, since it, alone among the methods identified, attempted to capture what is ultimately of value, namely improved health. None of the studies reviewed made any attempt to capture the intrinsic value of new knowledge (the value of discovery for its own sake).

In what follows, we discuss in more detail two approaches to evaluating the impact of health research. First we examine those studies that have attempted to estimate the value to society of the health gains attributable to health research (the last of the four categories identified by Buxton and colleagues and called the return on investment approach). Second, we review an alternative method developed by Buxton and Hanney (1996), called the payback approach, which provides a comprehensive framework for capturing all of the instrumental impacts of health research. The payback approach was not included in Buxton and colleagues' 2004 review since its scope extends beyond economic considerations.

2.2.1 The Return on Investment Approach

Three efforts have been made to quantify in monetary terms the health improvements attributable to health research. The first effort was that reported in a series of papers commissioned by the Lasker Foundation in the USA (Hatfield et al 2000). The approach involved the work of several economists, first to quantify the improvements in health that had been experienced in the USA in the twenty years since 1970, second to ascribe a dollar value to the improvements in life expectancy and quality of life using standard economic methods, and third, to attribute an acceptable proportion of this increased value to spending on all health research in order to estimate the return on investment (Murphy & Topel 2003a). Cutler and Kadiyala (2003) documented the substantial improvements made in life expectancy and quality of life in the USA since 1970, much of which has come from reductions in mortality associated with cardiovascular disease. They suggested that at least one-third of these gains in life expectancy and possibly as much as one-half could be attributed to health research (Cutler & Kadiyala 2003) providing a conservative estimate of the benefits of all health research. The health gains so estimated were then valued in monetary terms using contingent valuation methods from economics (see box). After reviewing the literature the value of each fatality avoided (and by extension the value of commensurate improvements in quality of life) was set at $3 million (Murphy & Topel, 2003a).

The results of the analysis were described as 'mind-boggling' by those who sought to summarize the project (Hatfield et al 2000). The improvements in health in the USA over the period 1970 to 1990 were valued at $57 trillion or $1.5 trillion per year. If, as suggested, one third of this improvement was due to health research, then the social value of this improvement in health ($500 billion per year) would cover the costs of the research twenty times over.

Putting a monetary value on people's lives: the contingent valuation approach

Life is priceless, but this does not mean that economists cannot put a monetary value on it! Actually what is valued is a small reduction in the risk of death, from which the value of a 'statistical life' can be computed (Jones-Lee 1976). This is what the contingent valuation approach seeks to do. In economics, the value of something is revealed in what one is prepared to sacrifice (or pay) in order to get it. Contingent valuation methods ask people what is the most that they would pay for something that reduced their risk of death by, say, 1%. It might be a smoke alarm installed in their house or an airbag installed in their car, or the introduction of speed cameras along a busy highway. Let's say we asked the same question of 100 people: 'what is the most that you would be willing to pay for a smoke detector that would reduce your chances of dying in a house fire next year by 1%?' If the smoke detector is installed in 100 houses then, on average, there would be one fewer fatality amongst this group of people. In this way, the intervention can be said to save one 'statistical life' each year. The value of this statistical life can then be inferred from the total amount that the group is willing to pay for the safety improvement. An alternate approach is to look to see how much monetary compensation a person requires to make them willing to face an increased risk of injury or death, such as that reflected in the wage premiums paid to workers in risky occupations (Viscusi & Aldy 2003).

Several studies have sought to quantify the value of a statistical life in this way. After reviewing the literature, the economists commissioned to estimate the return on health investment in the USA came up with a range of estimates of $3 million to $7 million per statistical life (Murphy & Topel 2003a). To be conservative, they used the lower figure.

The contingent valuation approach is not without its critics and it is fair to say that there is still much work to be done to improve the reliability and validity of the methods, and their acceptability (Klose 1999). Such methods are used routinely by government transport departments around the world however to assess the health impacts of transport infrastructure projects and road safety measures. They have been used in health economics evaluations as well, but not as widely (Diener et al 1998).

The same method was subsequently adapted and used to value the returns on investment to all health research in Australia first in 2003 (Access Economics) and secondly in a follow up study published in 2008 (Access Economics). Here it was assumed that 50% of all improvements in life expectancy and quality of life could be attributed to health research, and based on Australia's share of all health-related citations that between 2.4% and 3% of this could be attributed specifically to Australian research. Even with this adjustment for the share of the world's health research emanating from Australia, the results look impressive. Returns on investment to health research were estimated to be 1.4 in 2003 and 1.17 in 2008. That is every dollar invested in research yielded net benefits to society (that is benefits over and above the costs) of at least $1.17 and perhaps as much as $1.40. The reduction in the rate of return between 2003 and 2008 is attributed by the authors of the report to the fact that research spending had increased in Australia over this period at a faster rate than the expected increases in health that will follow.

The ROI method is obviously rather crude and the results depend to some extent on assumptions that are readily criticized (Siegler et al 2003). The results of the Australian exercise do remain fairly robust to changes in at least some of the underlying assumptions (for example, the return on investment remains substantially positive even if only 30% of all health gains can be attributed to health research rather than

50% as in the base case). However, only a limited range of assumptions were examined in this way and this limits the value of the sensitivity analysis. The main reason why the results are robust is probably the high value attributed to each year of healthy life gained. This drowns out most other considerations and tends to dominate this sort of analysis. Indeed, those involved in the original American study concluded that health research was bound to be a good investment even if it was responsible for only a very modest improvement in life expectancy, because the value of the health gains were so substantial relative to the costs of the research (Hatfield et al 2000).

Perhaps the most critical issue with the ROI approach is the accuracy with which improvements in health over time can be attributed to specific investments in research. One needs to specify the time lag between spending on health research and subsequent health improvement so that the costs of the research can be correctly identified, and to separate the effects of the research from the myriad of other social, economic, demographic and political forces that contribute to increasing health status over time. The failure to deal adequately with this issue leads to perverse results. In the 2008 Australian study, the authors found that health outcomes were getting worse for three classes of disease: endocrine and metabolic disorders (predominantly diabetes); mental disorders, nervous system and sense disorders; and musculoskeletal diseases. The logic of the ROI method implies that the return on investment in these areas has been negative, but not according to the authors of the report. Rather this meant that not enough had been spent on research to counteract the increase in prevalence of these diseases from other causes (Access Economics 2008). Thus, where health gains were positive, the authors concluded that research was effective and the return on investment was positive. Where health gains were negative, then research was still effective but spending was inadequate. In either case, the inference one draws from the authors' conclusions is that spending on research has positive value and needs to be increased. It is not clear therefore how one might ever conclude that the return on investment in health research was inadequate and that too much was being spent on research.

One other notable omission from both the original analysis and the Australian studies was any consideration of the costs of the health care required to realize the potential gains from research (that is, costs to translate the new knowledge into practice). The summary of the American work acknowledged that this would reduce the return on investment, but there was no rigorous attempt to quantify by how much. The impact of health spending on returns to investment in research has since been estimated by two of the economists who led the original American effort (Murphy & Topel 2003b). As expected, the inclusion of health care costs does indeed reduce the return on investment and while for men, the return remains positive, for older women, the costs exceed the benefits. Women have not benefitted as much as men from the reduction in cardiovascular mortality but they bear an equi-proportional share of the costs of the research. The real problem, however, according to the authors of the report, is distortions in medical decision making, caused by perverse incentives in the US health care system that lead to the overuse of new health technologies. The authors acknowledge that such distortions could mean that future returns on investment in medical research need not be positive.

Factoring in the cost of implementing research findings is also important for prioritizing research and for determining the return on investment to different sorts of research activity. Cutler and Kadiyala (2003), two of the researchers involved in assessing how much of the improvement in health could be attributed to health research in the original US estimate of return on investment, claimed that one third of the reduction in cardiovascular mortality was due to invasive therapies such as coronary artery bypass grafting, one third was due to pharmacological advances and one third to behaviour change especially with respect to smoking. Success in each of these is underpinned *among other things* by health

research, and in respect to behaviour change especially by population and public health research, but primarily because population health strategies rarely cost as much as invasive therapy or pharmaceuticals, they offer a greater return on investment. Cutler and Kadiyala (2003) suggest the return on population and public health research in the USA since 1970 has been 30:1 compared with 4:1 for biomedical research.

However, the real merit of the ROI approach lies not in the questionable estimates that it provides but rather in the discussions that will be associated with any attempt to apportion changes in health status to each of the various determinants. For example, to our eyes, the real benefit of the US and Australian studies is how they each demonstrate the intersections between, and the complementarity of, research across the spectrum from the biomedical and clinical sciences through to health services, and population and public health research. The tortured conclusion reached by the authors of the Australian follow-up study in the face of worsening health burden for some diseases suggests the need for increases in population and public health research designed to reduce the incidence of disease, not necessarily more biomedical or clinical research to improve treatment of the consequences. Similarly, the effect of health care costs on the returns to US spending on research suggests that more health services research aimed at improving medical and health care decision making, perhaps through better information systems or more appropriate incentive structures, could improve rates of return to research in the biomedical and clinical areas.

There will undoubtedly be disagreement about how much health change can be rightfully attributed to different types of health research, but within the bounds of this disagreement the discussion will have one of three outcomes. Either we will find positive returns on investment, no matter how pessimistic the attribution of health gain to research, or we will find a negative return on investment no matter how optimistic we are in attributing health gains to research. More likely however, we will find that our estimates of the return on investment are sensitive to the assumptions we make about the proportion of health improvement that is caused by health research. In this case, the ROI exercise will have served a purpose even if it only shows the futility of over-simplistic efforts to quantify the value of health research.

2.2.2 The Payback Approach

The second method we consider is the 'payback' approach developed by Martin Buxton and Steve Hanney in the UK (Buxton & Hanney 1996). Unlike the return on investment approach discussed above, the payback approach involves a comprehensive categorization of all of the potential benefits of health research. It also includes a logic model that sets out the linkages between investment in health and medical research and each of these potential impacts. As such, it encompasses the economic benefits identified in Buxton and colleagues' original review of the literature and more besides (Buxton et al 2004). The method also looks beyond just the generation of new knowledge and covers the whole research process, from initial choice of research topic through to dissemination of the results and the incorporation of research findings into policy and practice.

The original framework has been revised over the years and applied in several different contexts. It has been used to evaluate clinical research, such as the use of antenatal corticosteroids to prevent neonatal respiratory distress syndrome (Hanney et al 2005), to assess the value of disease-specific programmes of research in arthritis (Hanney et al 2004) and diabetes (Hanney et al 2006) and to evaluate the impact of national research programmes such as the UK National Health Service's research and development programme (Buxton & Hanney 1998; Soper & Hanney 2007).

The payback approach specifies five categories of 'impact': knowledge production; research targeting, capacity building and absorption; informing policy and product development; health benefits; and broader economic benefits (see text box for descriptions of each category). Unlike the 'return on investment' method, application of the payback approach does not allow one to calculate a single, summary measure of the value of research. Rather, a more detailed picture is developed with counts of outputs that can be easily quantified (e.g., publications, PhD projects, citations) as well as a narrative description of other impacts (e.g., reference to research findings in the development of clinical guidelines or in health policies).

The attribution problem discussed above with respect to the ROI approach is not so much dealt with as 'accommodated', in the sense that no effort is made to quantify precisely the contribution that research makes to higher level outcomes such as health gains. Instead, the narrative describes the connections and associations, and provides a summary of the various impacts of the research to inform one's judgment about its value.

Categories of Payback to Health Research (adapted from Hanney et al 2004)

Knowledge Production: Described by Buxton and colleagues as the first product of research, this is traditionally measured by journal publications, citation analyses, the impact factor of the journal in which the research appears and the relevance of the journals to the intended audiences.

Research targeting, capacity building and absorption: This category refers to the advantages that previous research confers on future research activity. Targeting refers to the benefits that accrue to later researchers who build upon previous knowledge. It refers both to advancing knowledge and to increasing the amount of research in an area, such as when good researchers are able to lever additional resources for research into a specific field. Capacity building refers both to researcher-training (for example research-based degrees aligned with particular research activities), as well as to receptor capacity such as the research literacy of research-users that facilitates the uptake of new knowledge.

Informing policy and product development: This category refers to the use of research to develop policies or new products. It is acknowledged that the policies have to be implemented and new products brought to market before their value can be realized.

Health gain: This is regarded as the 'real' payback to health research. The category includes health improvement (measured by the reduction in potential years of life lost or the number of quality-adjusted life-years gained for example), reductions in the cost of health care; improvements in the quality of service delivery, and; reduction in health inequalities (equity considerations).

Broader economic benefits: Included here are the economic benefits of having a healthier workforce (reduced absenteeism leading to increased production) as well as benefits to the national economy arising from the commercialization of the products of research.

2.3   Reflections on a Framework for Canada

The payback approach resonated better than the return on investment approach with the expert participants attending the previously mentioned CIHR sponsored workshop, though some modifications were recommended to adapt the approach to Canadian uses (Canadian Institutes for Health Research 2005). All economic impacts were consolidated into a single category. This involved taking commercialization out of 'informing policy and product development', taking savings in health care expenditure out of 'health gain' and including both in a re-labelled 'economic benefits' category. The new categories in the amended version were: knowledge production; research targeting and capacity; informing policy; health and health sector benefits; and economic benefits. Several audiences for the information generated by the payback approach were identified, and these included: the higher education sector; health professionals and administrators; the business sector; government and the wider society. It was suggested that the reporting of research impacts should be tailored differently to the needs of each these groups. Finally the panel of experts proposed a number of indicators of research impact and identified sources of information for these. We discuss these indicators and suggest additional metrics in Section 4.

3. Population and Public Health Research

3.1   Scope of Population and Public Health Research

We turn now to discuss what is meant by population and public health research (PPH), since one issue we must confront in considering what frameworks might prove most useful for assessing the social value of PPH research is in deciding what this covers and how much has been spent on PPH research activities with what ends in mind.

Public health has been defined as 'the science and art of preventing disease, prolonging life and promoting health through the organized efforts and informed choices of society, organizations, public and private, communities and individuals' (Last 2001). *Population health* is regarded as an approach to health that must aim to improve the health of the entire population and to reduce health inequalities among population groups by tackling the root causes – the social, cultural and environmental determinants of health across the lifespan (Strategic Policy Directorate 2001). The population health approach is often associated with the work of Geoffrey Rose (1992) who argued that shifting the whole distribution of risk may improve aggregate health more than any single targeted high risk strategy. The argument is a contested one, but one which is ultimately an empirical question (Charlton 1995; Manuel et al 2006). Moreover, some authors have argued that such an approach can inadvertently increase inequalities thereby calling on the complementary use of both a population health and a vulnerable populations approach (Frohlich & Potvin 2008).

CIHR categorizes its research activities into four themes or 'pillars'. These are: (i) biomedical; (ii) clinical; (iii) health systems and services; and, (iv) social, cultural, environmental and population health. Pillar IV was initially just defined as the social, cultural and environmental influences on health. The addition of 'population health' to the labelling of the pillar IV category came later. Despite this, the theme is often identified just with population health.

Pillar IV research is now defined as:

> 'Research with the goal of improving the health of the Canadian population, or of defined sub-populations, through a better understanding of the ways in which social, cultural, environmental, occupational and economic factors determine health status ' (Canadian Institutes of Health Research 2008a)

Despite the 'population health' appendage, pillar IV research is not identified solely with the Institute of Population and Public Health (one of the 13 virtual institutes through which the CIHR seeks to integrate research across the four pillars or themes). Instead, most of the 13 institutes have responsibility for research across the four themes (though as we will see shortly the extent to which this is reflected in resource allocation varies markedly among institutes). Neither, as we will see, is pillar IV research inclusive of all research relevant to population and public health.

The ambit of the Institute of Population and Public Health (IPPH) encompasses pillar IV research but extends beyond it to emphasize the *interactions* between the various determinants of health, including the biological. Alone among similar research funding bodies around the world, the IPPH explicitly combines the population health approach with a concern for public health policy and practice. Specifically the IPPH:

> 'supports research into the complex interactions (biological, social, cultural, environmental) which determine the health of individuals, communities, and global populations; and into the application of that knowledge to improve the health of both populations and individuals' (Institute of Population and Public Health 2008a).

Scope of IPPH Research

- health promotion policies and programs (individual, community, and population based); related health outcomes research
- health determinants - to elucidate the multi-dimensional factors that affect the health of populations and lead to a differential prevalence in health outcomes
- identification of health advantage and health risk factors related to the interaction of environments (cultural, social, psychological, behavioural, physical, genetic)
- disease, injury and disability prevention strategies at the individual and population levels; identification and study of special populations (e.g., rural populations)
- environment and health (e.g., radiation, contaminants, ecosystem and health, air quality)
- socio-economic and cultural determinants of health (e.g., poverty, social status, access to public health services, literacy, community characteristics)
- public health and community health issues - surveillance, monitoring, information and data, laboratory studies and their implications for public health (e.g., safe water)
- workplace and occupational health research including physical, chemical, biological and organizational factors in the workplace
- health policy formation at community, regional, provincial, national and international levels; relation to health outcomes
- basic methodology development (e.g., epidemiology, biostatistics, survey development, surveillance tools, tools for risk evaluation, risk perception, modeling complex interactions)
- multiple interventions research to determine the best combination of interventions, providers, and conditions to improve the health of populations and reduce inequalities in health
- underlying mechanisms through which social and physical environments influence human biology
- development and implementation of health technologies and tools to support public health decision-making (e.g. surveillance technologies, detection devices, database design)
- ethical issues related to population health (e.g. poverty, exposure to hazards)

Both the CIHR definition of pillar IV research and the scope of the research supported by IPPH make reference to the link between better understanding of the determinants of health and efforts to improve population health. The scope of both therefore includes basic (social science) research to illuminate pathways and relationships that generate health in populations as well as research into the design, implementation and evaluation of interventions aimed at disease prevention and protecting and improving health. These interventions include *policies*, such as bans on tobacco use in public places or the introduction of voluntary codes of practice on food product labelling, and *programmes* such as social marketing campaigns to influence healthy eating and active living. It should be noted that it is common, but incorrect, to refer to all PPH research as 'applied research' as if it necessarily involves using knowledge obtained though pillars I and II for example, and 'applying it' to whole populations. While there is an element of this (for example where knowledge from the laboratory is used to generate whole population interventions, such as vaccination, for example) such framing takes away the recognition that a vast amount of 'basic' research also happens at the population level. This 'basic' research serves to illuminate pathways and relationships that generate health in populations, drawing on disciplines such as anthropology, communication, economics, epidemiology, geography, mathematics, political science, psychology and sociology.

To complicate matters further, not all research relevant to public health fits entirely under pillar IV or the scope of the IPPH. Vaccination for example is a mainstay of public health policy but the research needed to support effective vaccination policy spans all four CIHR pillars (see box).

Public Health Research Supported by all Four Pillars of CIHR

Historically, vaccination against infectious diseases has contributed to substantial reductions in mortality and morbidity, reductions in health care costs and improvements in economic production. Once common, infectious disease now contributes very little to the overall burden of ill health in Canada (with some exceptions). Returns on investment in vaccination typically far exceed the costs.

Acclaimed as a public health success story, the research underpinning effective vaccination programmes spans the four CIHR pillars. Biomedical research (pillar I) is required to identify and isolate protective antigens, to create new vector systems and to develop and evaluate adjuvant systems. Clinical research (pillar II) examines the safety, immunologic response and efficacy of the vaccine. Health services research (pillar III) might then examine the cost-effectiveness of alternative delivery modes, looking to see whether the vaccine is best administered by primary care physicians or nurse practitioners, in fixed or mobile clinics or in schools. Finally, population health research (pillar IV) would include the original epidemiological evidence pointing to the potential value of a vaccine, descriptive research identifying the social and cultural barriers and facilitators to uptake of the vaccine, and evaluative research to assess the cost-effectiveness of interventions designed to increase vaccination rates.

Each contribution of research from each of the four pillars is essential in understanding and maximizing the impact of vaccination. Each piece is necessary, but no one piece is sufficient. Many of the benefits of vaccination (and therefore the research that supports it) especially the health gain that results, are easy to quantify, but efforts to apportion this benefit among the four pillars in order to isolate the return on investment in any one programme of research would be difficult if not meaningless.

The IPPH's ambit fills a gap between the CIHR pillars by supporting research on the interactions between individual and other determinants of health. It also aligns better with definitions of public health and the population health approach than does CIHR's definition of pillar IV research. We have not examined how the other CIHR institutes define their scope or operationalize their pillar IV responsibilities, but it is likely that they too will be addressing gaps at the interface between the CIHR pillars and in so doing support activities that might rightfully be included within population and public health research.

3.2   How Much Does CIHR Invest in Population and Public Health Research?

Based on what is published on the CIHR website, an indication of the allocation of research funds across each of the four pillars since the inception of the CIHR is shown in Figure 1.

Figure 1:   CIHR Grants and Awards Funding by Research Theme 1999-2007
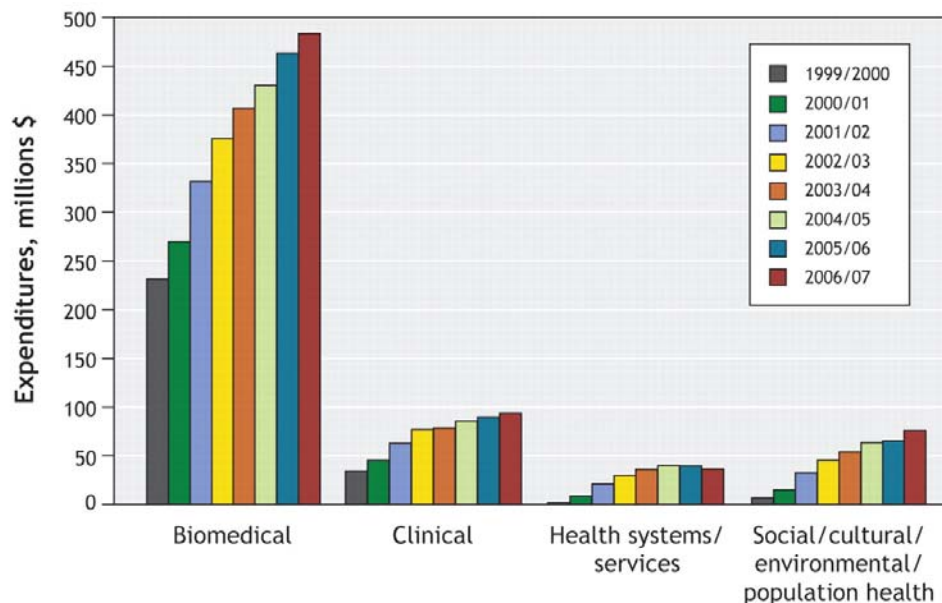
Pillar IV research has increased more than ten-fold over the period, albeit from a small base. In 1999-2000, spending on pillar IV research amounted to less than $7 million. In 2005-2006, it totalled more than $70 million. Over the same period, the total CIHR budget more than doubled. These numbers need to be interpreted with some caution however. When researchers apply for funding they are asked to self-designate into which CIHR pillar(s) their research falls. This information was not previously validated, and an analysis conducted by IPPH demonstrated how self-designation lacked the sensitivity and specificity required to confirm with confidence what proportion of CIHR funding truly goes to the pillar IV research community (Institute of Population and Public Health 2004). For example, the IPPH analysis found that use of epidemiological or social science methods was frequently categorized as pillar IV research irrespective of whether the study was focussed on population health questions.

The problems with self-identification also apply to the alignment between research funds and a given CIHR institute. Subject to the same caveat as discussed above therefore, pillar IV research is also spread unevenly across the 13 institutes (Table 1). The table shows cumulative spending by CIHR since 1999-2000 to May 2008 (the figures are derived from those published on the CIHR web-based funding data-base). Of the $6.25 billion spent by CIHR over this period, $466 million (7.4%) was attributed to pillar IV research. One third of this was identified by the applicants as relevant to the Institute of Population and Public Health. Other big 'contributors' in dollar terms to population and public health research were the Institute of Human Development, Child and Youth Health, the Institute of Aboriginal Peoples' Health and the Institute of Gender and Health. Note however, that the institutes are not responsible for allocating

all of these funds. Rather, the attribution shown in the table is largely the result of self-identification by research applicants. (As mentioned above, CIHR does now validate some of the grants using subject experts, but this practice has only been implemented in the past few years).

TABLE 1: Relationship between Pillar IV Research and CIHR Institute

INSTITUTE	PILLAR IV RESEARCH		ALL PILLARS		PILLAR IV AS % OF ALL RESEARCH	
	Grants	Funds	Grants	Funds	Grants	Funds
Aboriginal Peoples' Health	167	$55,426,848	232	$85,405,958	72.0	64.9
Ageing	117	$29,914,986	737	$157,897,352	15.9	18.9
Cancer Research	51	$12,883,834	1741	$479,158,352	2.9	2.7
Circulatory & respiratory health	33	$8,600,581	1855	$644,719,181	1.8	1.3
Gender and Health	180	$35,151,412	385	$89,778,572	46.8	39.2
Genetics	43	$10,181,726	1557	$599,776,130	2.8	1.7
Health Services & Policy Research	68	$12,601,653	994	$190,699,577	6.8	6.6
Human Development, Child & Youth Health	215	$56,637,716	1033	$362,784,390	20.8	15.6
Infection and Immunity	130	$30,705,413	1982	$613,154,297	6.6	5.0
Musculoskeletal Health and Arthritis	43	$14,049,243	970	$286,633,092	4.4	4.9
Neurosciences, Mental Health & Addiction	137	$26,140,812	2695	$828,853,602	5.1	3.2
Nutrition, Metabolism & Diabetes	95	$12,297,135	1238	$373,076,543	7.7	3.3
Population and Public Health	716	$153,414,218	967	$203,515,611	74.0	75.4
Not Specified	37	$8,706,115	7239	$1,349,159,766	0.5	0.6
Totals	**2032**	**$466,711,692**	**23625**	**$6,264,612,423**	**8.6**	**7.4**

Most of CIHR's research money (70%) is allocated independently of the institutes through the peer-review committees in the open competition (the investigator-initiated pool of funding). In this instance researchers formulate the ideas, develop the proposals and submit the applications. Only 30% of the CIHR budget is allocated to strategic initiatives, of which approximately 12% falls directly within the purview of the institutes. Strategic initiatives are developed by the 13 institutes in response to priorities, and usually stipulate additional requirements including specific research themes to be addressed.

CIHR has introduced measures to validate the assignment of research proposals to pillars, but limitations of relying on investigator self-designation of research pillar and research institute remain. There is a need therefore to implement a better system of data collection and analysis to improve the monitoring of what proportion of funds are invested by pillar, as well as to develop a more robust research classification system that is more in line with CIHR's broad vision for health research. Such a system would not only enable the better mapping of the full, interdisciplinary scope of health research, but also reduce reliance on key word searches and improve the effectiveness and efficiency of current analyzes of research funding by pillar. Many of these activities are fortunately in progress.

3.3 What Sorts of PPH Research Does CIHR Invest In?

Not all of the research dollars attributed here to population and public health are spent on the generation of new knowledge. The objective of the CIHR, as spelled out in the Act by which the institutes were established is:

> 'to excel, according to internationally accepted standards of scientific excellence, in the creation of new knowledge and its translation into improved health for Canadians, more effective health services and products and a strengthened Canadian health care system' (emphasis added). (Canadian institutes of Health Research 2008c)

Thus, in addition to generating new knowledge, CIHR has a responsibility to ensure that this knowledge is taken up by policy makers and practitioners and implemented for the good of all Canadians.

It was also recognized early in the life of CIHR, that in the respect of PPH research the institutes would need to invest first in capacity building, both on the research side (increasing the supply of competent population and public health researchers) and on the demand side, building receptor capacity among public health agencies and other relevant agencies. In response to criticism that health services research and population and public health research were under-funded, Alan Bernstein, the first President of CIHR, reported that the 'the slow growth rate in those areas was largely a function of a lack of critical mass within those sectors during the agency's formative years.' (Kondro 2007).

In practice, most CIHR funding initiatives, including grants for investigator-led research, span multiple objectives, including research capacity building, knowledge generation and knowledge translation. Population and public health research is no exception. For example, all operating grants, salary awards, fellowships and scholarships include a proposal and plan for new original research. A detailed knowledge translation plan is also important to the success of investigator-led proposals. Moreover, the institutes often issue 'Requests for Applications' that delineate specific objectives and other requirements. For example, the newly established Chairs in Applied Public Health programme, (a joint program of IPPH involving the Public Health Agency of Canada and other partners) explicitly combines knowledge generation, capacity building and knowledge exchange. While the substantive research

content of these chairs was determined by the applicants (and so resembles somewhat investigator-led research) the program required a focus on intervention research and set other boundaries (e.g., education and mentoring, links to public health organizations, and knowledge translation requirements) in line with the strategic intent of the initiative: namely to build the research base underpinning public health policy and practice in Canada.

Given the integrated focus of much CIHR funded research, it is difficult to attribute research funding specifically to knowledge generation, knowledge translation and/or capacity building. We nonetheless try to attribute pillar IV funding in this way (Table 2). The allocation of funds is based on our assessment of the primary objective of the funding stream. (An alternative approach, apportioning the budget for each award to each category would be a resource intensive exercise and of dubious added-value). All operational grants awarded in open competition or as part of the randomized control trials competition were classified as 'investigator-led research'. Grants awarded under named competitions, such as the Advancing Theories, Frameworks, Methods and Measurement in Health Services and Policy and Population and Public Health programmes, were labelled as 'strategic research' because such competitions were usually led by one or more institutes. All personal awards such as new investigator awards, scholarships and fellowships, and the Chair's programme discussed above were regarded as 'research capacity building'. (Note that the research content of these awards is determined by the investigator, and so they could also be classified as investigator-led knowledge generation. We regard them as capacity building, especially in a relatively less well-supported field as population and public health. However, the primary intent of these awards is to ensure sufficient supply of competent, well-trained, well-supported researchers. Moreover, the salary award typically covers the costs of the investigator and perhaps an otherwise uncommitted research allowance. (Separate operating funds usually have to be acquired to support the costs of the research.) For the same reason, the funding for the seven Centres for Research Development was also regarded as infrastructure to strengthen research capacity in population and public health and its effective use, and which included as part of the budget funds for pilot research projects.

TABLE 2: Allocation of Pillar IV Research Funds by Type of Award (2000-2008)

Type of Award	No of Awards	Total Value of Award	Funding as % of Total
Investigator-led research	774	$232,044,442	50%
Strategic research	360	$57,595,920	12%
Research capacity building grants	818	$155,484,232	33%
Knowledge exchange	81	$5,761,447	1%
Administration	2	$16,154,818	3%
TOTALS	**2035**	**$467,040,859**	**100%**

By this estimate, 50% of pillar IV research is investigator-led: considerably less than the 70% average for CIHR as a whole (though note, our method of assigning grants to categories is not the same). A small proportion of funding (just 1%) directly supported knowledge translation (KT) activities (these are the KT workshops that CIHR fund) though many of the strategic research grants require an integrated knowledge translation component and so this figure underestimates the resources devoted to knowledge exchange. As expected given concerns about the lack of critical mass in PPH research, one

third of pillar IV funding has supported capacity building through salaries and awards to investigators and trainees, training grants, network building and the Centres for Research Development programme.

There is undoubtedly some error in this attribution (for example, to the extent that salary awards and training awards include an operational component we have underestimated investigator-led research) but it does reflect one of the priorities of the IPPH, namely to build a constituency of researchers capable of advancing population and public health research. This has been the major obstacle preventing Canada from capitalizing on the lead it held in this field during the 1980s and 1990s (see box).

'Canadian researchers have been the most influential force in the world in demonstrating the importance of population determinants of health, but research to realize these ideas has generally been conducted elsewhere in the world', says Dr. Leonard Syme, Professor Emeritus of Epidemiology, School of Public Health, University of California at Berkeley and IPPH Advisory Board Member. The reasons for this are two-fold – there are actually very few population health researchers in Canada and there have been very limited funds to support their work.

Institute of Population and Public Health 2008b

Looking ahead therefore, one indicator of the success of this strategy ought then to be an increase in applications for funding from the open competition from population and public health researchers as the cohort of trainees who have benefitted from previous capacity-building efforts 'graduate' to become lead investigators in their own right.

Case Study: A Public Health Success - Comprehensive Tobacco Cessation Programs

In July 2004, the city of Saskatoon enacted a ban on smoking in public places. Research funded by a coalition of agencies including the CIHR has recently reported on the impact of the ban (Lemstra et al 2008). The incidence of acute myocardial infarction (AMI) fell from 176 cases / 100,000 people in the years before the ban to 152 / 100,000 population in the year afterwards. There were 32 fewer hospital admissions for AMI, smoking prevalence fell from 24.1% to 18.2% in the city whilst remaining relatively constant in Saskatchewan and the rest of the country, and people who continued to use tobacco smoked 6.6 fewer cigarettes each day (Lemstra et al 2008).

Saskatoon was just another in a long line of jurisdictions to take action on tobacco following the lead set by California in 1988 with its passing of Proposition 99. This increased taxes on tobacco products substantially and allocated 20% of the revenues raised to implement a comprehensive tobacco control program. In the eight years following the implementation of Proposition 99 there was 14,000 fewer heart attacks and strokes, 10,800 fewer low birthweight births and 2,500 fewer deaths. Savings to the health care system amounted to more than $500 million (Lightwood and Glantz 1997; Lightwood *et al.*, 1999). California's policy lead was closely followed by Massachussets, Arizona and Oregon. Now more than 75 countries around the world and 25 US states have imposed bans on smoking (Schmidt 2007). The reduction in premature deaths and illness relating to tobacco use is one of the great public health success stories of the twentieth century.

Proposition 99 was the culmination of many years of local advocacy and community organisation around non-smokers' rights. Population and Public Health research, primarily documenting the health effects of tobacco use and exposure to second-hand smoke, was necessary, but played only a small part in the legislation's passage (Glantz & Balbach 2000). PPH research carried out since the implementation of Proposition 99 has been more instrumental in fuelling the spread of comprehensive tobacco control strategies around the world. This research has continued to document the increased risk of disease associated with exposure to second hand smoke thus making tobacco control a social issue (e.g., Ducatman & McLellan 2000), the costs associated with tobacco use (e.g., Warner et al 1999), the effectiveness of interventions to reduce uptake and consumption (e.g., Fichtenberg & Glantz 2002; Siegel 2008), the reduction in exposure to second hand smoke following bans on smoking in public places (e.g., Galan et al 2007; Stark et al 2007), and the improvement in health outcomes that have followed (e.g., Eisner et al 1998; Eagan et al 2006; Goodman et al 2007).

If further evidence of the value of this research is needed it can be seen in the extraordinary lengths to which the tobacco industry will go to discredit the science (Chapman, 1997; Samet & Burke, 2001; Trotter & Chapman 2003).

It is difficult to attribute the huge gains that have been made with respect to tobacco control to particular pieces of research evidence however. With each new policy and each new study, the evidence base increases and policy change becomes easier to instigate and sustain.

Case Study: Population and Public Health Research Contributes to Halving of Infant Deaths

In 1993, in response to emerging evidence about the risk factors associated with sudden infant death syndrome (SIDS), the Canadian Paediatric Society recommended that all infants be placed on their back to sleep (Injury Prevention Committee 1996). In so doing Canada joined the USA, the UK, Australia, New Zealand and others in an evidence-base campaign (aptly named the 'Back to Sleep' campaign) to reduce sudden infant deaths. Prior to the campaign more than one baby in every one-thousand born alive in Canada died from SIDS. In the years after the campaign, the rate of SIDS has fallen by nearly one-half (Rusen et al 2004). Similar reductions have been seen around the world (Davidson-Rada et al 1995; Dwyer et al 1995; Gilman et al 1995; Mitchell et al 2007; Wigfield et al 1992).

The evidence that supported the back to sleep recommendations around the world was solidly 'pillar IV', population and public health research. This involved the painstaking examination of the interactions among risk factors and differences in SIDS rates from population-level studies in several countries to identify putative pathways and possible entry points for intervention, supported by further population-level studies to evaluate the impact of new preventive policies, based on that evidence. No one study was critical in itself, but all were instrumental in informing and changing policy. It was the accumulating weight of evidence reaching a tipping point, rather than the insights offered by any one study that changed practice and saved thousands of young lives.

The example illustrates three points. The first is just how valuable population and public health research can be. In Canada alone, more than 200 young lives are now saved each year, most of which can be attributed to the increase in the proportion of infants being placed on their back to sleep (Rusen et al 2004). The second insight is just how difficult it will be to try and attribute a portion of this success to one country's investment in research. Finally, since the critical mass dedicated to any one field of study may be too limited in one country, contribution to the global evidentiary base has to be a key objective of research funding.

---

4. Measuring the Social Value of Population and Public Health Research

4.1   Which of the Frameworks Best Suits Canada?

We turn now to consider how best to evaluate the impact of CIHR's investment in PPH research. Our discussion of CIHR spending on PPH research identified three challenges that we face when thinking about how best to measure the social value of this investment. The first is in identifying what counts as PPH research so that we may quantify the amount that has been spent. The second is in being sensitive to the multiple objectives of this spending and the implications this has for identifying and measuring impact, particularly as different objectives have taken priority at different times. The third lies in disentangling the contribution that CIHR has made from the contribution of its partner agencies, where there is value-added from the collaboration.

The SIDS example, with which we opened this report, the vaccination case featured in Section 3, and our discussion of the limitations of the return on investment approach also illustrate a fourth challenge, one of attribution. The contribution that research makes to a successful vaccination campaign spans all four CIHR pillars. Each pillar's contribution is necessary, none are entirely sufficient to ensure that a laboratory-based discovery is translated into effective public health policy. With SIDS, the problem manifests itself slightly differently. The research underpinning the successful 'Back to Sleep' campaigns of the early 1990s was solidly 'pillar IV', but it was the weight of evidence amassed over many years from several different countries each pointing to the same conclusion that prompted action around the

world. Again, all of the studies were necessary but no one study was critical in changing policy and practice. The problem of attribution here lies in deciding how much of any improvement in population health is down to CIHR-funded research.

Given these complexities, the payback approach of Buxton, Hanney and colleagues, as amended by the experts attending the CIHR organized discussions on developing a framework to measure the impact of health research, offers much promise. We return to discuss this point in more detail shortly, but we should not lose sight of the benefits offered by a rapid appraisal of the return on investment to PPH research of the sort originally commissioned by the Lasker Foundation in the USA. Such an appraisal is relatively easy to complete. National agencies such as Statistics Canada and Health Canada have the necessary data on improvements in health and on total spending on health research. In addition, we need information on Canada's share of health-related citations. If this is not readily available then it will need to be compiled anyway as it will be a vital component of any assessment of the social value of health research. It is needed here to estimate the proportion of the total improvement in health attributable to Canadian research. The result of the exercise will be crude but it will provide some indication of an order of magnitude, which if the results turn out positive will be reassuring information politically. The results could also be readily compared with the outcomes of the US and Australian efforts, thereby raising legitimate questions about the relative performance of Canada's population and public health research community.

The return on investment approach is unlikely to provide a universally acceptable estimate of the value of particular programmes or research themes, not least because of the difficulties associated with gaining agreement on the proportion of health gains that should be rightfully attributed to which research programme or theme: something that is further complicated by the increasing recognition that finding solutions to complex health problems requires interdisciplinary and cross-pillar approaches and at times international collaborations to facilitate cross-national comparisons. The discussions around this point would be interesting to have, however and valuable even if they led only to agreement that it is difficult if not impossible to partition research outputs in this way.

The payback approach is more flexible than the return on investment approach and so may be better adapted to the challenges we noted at the beginning of this section. It better accommodates the different stages of development of each of the CIHR themes since the five-fold categorization of types of impact enables one to match measures of success with the objectives of particular funding streams. Thus, for population and public health research, we can rightfully expect to see outcomes in the form of increased capacity, but are unlikely to see improvements in health outcomes or economic benefits, at least not yet given the considerable lag effects. There ought to be signs though of progress towards these outcomes, as measured by increases in citation rates of CIHR-funded research and the influence that this is having on statements about health policy (and potentially other sectors, depending on the research focus) and public health practice. The use of a logic model will also allow one to examine the complementary nature of research and perhaps also to tease out if only in crude terms, the value-added of the different types of research activity.

The payback approach is more time-intensive than the return on investment approach however, and is therefore a more costly exercise. It should be possible though to tailor how the payback approach is used, so that frequent assessments are made using a small number of indicators that are relatively easy to obtain (e.g., PhD completion rates, and citation rates), supported by occasional, more intensive

evaluations of a larger range of indicators that provide a more comprehensive picture of payback or a better assessment of payback in one or more fields.

4.2   How might the Payback Approach be Applied to PPH Research?

For the reasons discussed above, the payback approach resonated well with the expert panel convened by CIHR. The panel suggested a range of indicators that could be used to operationalize the approach (Table 3). Information on many of these indicators is collected routinely by CIHR as part of its Management Resources and Results Structure. Those marked 'N' are new measures suggested by the CIHR expert panel. Indicators in italics are new measures suggested by us. The original indicators suggested by the panel were limited in their scope (e.g., capacity building is limited only to the Canada Research Chairs (CRC) program), and sometimes rather vague (e.g., 'strategic research initiatives and their outcomes') and will need to be unpacked, but they provide a starting point. We discuss some of these indicators and those that we have added below.

4.2.1   Knowledge Production

In addition to the number of publications (a measure of research output) resulting from CIHR funded research, adding *citation counts* and evaluations of the *impact factors* of the journals in which the articles appear would provide an indication of research quality. *Relevance* of the journal to the target audience (i.e. the extent to which it is actually read by those responsible for implementing research findings) may also be a factor to consider and may conflict with the uncritical use of impact factors.

None of these bibliometric indicators is perfect. Their shortcomings are well-documented (Franks et al 2006), especially in respect of multidisciplinary work (Zitt 2005). We are not suggesting that they be used uncritically, or in isolation, but if the knowledge produced is of value to other researchers then we would expect to see this reflected in bibliometric measures (van Raan 2005). Of course, much-criticized articles also have high citation counts and so the content of the citations also needs to be considered.

Table 3   Indicators of Health Research (taken from CIHR Framework)

Knowledge Production	
1. Number of publications resulting from CIHR-supported research 2. High peer review rankings of results of CIHR-funded research 3. *Citation rates* 4. *Impact factors and relevance of journals to target audiences*	Bibliometric studies *Bibliometric studies* *Bibliometric and case studies*
Research Targeting and Capacity	
1. Extent to which Institutes have appropriately influenced the research, policy and/or practice agendas in their communities 2. Percentage of Research Chair holders attracted or retained in Canada due to the CRC program (*and Applied Chairs Program*) 3. *Postgraduates, PhD & Masters completions (rates)* 4. *Receptor capacity* 5. *Co-authorship and other evidence of interdisciplinary research* 6. *Co-citation network analysis (main path analysis)*	Evaluations every 3 – 5 years CRC database *IPPH database* *Case studies* *Bibliometric studies* *Bibliometric studies* *Bibliometric studies*
Informing Policy	
1. Number of public policies influenced by ethical legal social issue (ELSI) principles 2. Number of clinical practice guidelines by disease area influenced by CIHR funded research (N) 3. *Public policies influenced by PPH research*	Case studies Evaluations every 3 – 5 years *Case studies*
Health and Health Sector Benefits	
Public health: Strategic research initiatives and their outcomes (N) Health impacts: 1. Impact of health research on PYLL for target disease categories (e.g., cancer, circulatory disease) (N) 2. *Impact of health research on quality of life* 3. *Impact of health research on health inequalities*	Case studies Statistics Canada data Special studies to link to health research *Statistics Canada* *Statistics Canada*
Economic Impacts	
Commercialization: 1. Number and nature of patents, spin-off companies and licenses for intellectual property generated from CIHR funded research. 2. Income from intellectual property commercialization. 3. Case studies and follow-up surveys of commercial use of research funded by CIHR's Proof of Principle program.	Statistics Canada Statistics Canada Special studies
Cost savings: Estimates of the value of high impact innovations developed through health research in Canada (N) *(Need to include savings accruing to other sectors, e.g., education, welfare, criminal justice)*	Special studies
Human capital: 1. Reduction in productivity lost through illness or injury due to innovations from research (N) 2. *Increased productivity through improved human capital (e.g., child development)*	Collaborative studies with Health Canada and Statistics Canada

4.2.2   Research Targeting and Capacity

In addition to the CRC programme, the successes of each of the salary support programmes funded by CIHR need to be documented. These include the Applied Chairs in Public Health, the new investigator awards and the scholarships and fellowships programmes. CIHR is developing a range of performance

indicators for the Applied Chairs programme and these should feed into an evaluation of research impact. Other obvious measures of capacity include the number of post-doctoral fellows supported individually and via the Strategic Training Initiative in Health Research, and the number of new graduates (PhD and Masters) and undergraduates with training in or exposure to population and public health research.

It is also imperative to look beyond these direct capacity-building efforts to include any subsequent *cascade effects* of CIHR support. These are the spin-off effects that arise because CIHR supported researchers are able to lever additional resources for capacity building or because CIHR supported projects create new opportunities to add student projects incrementally. For example, the discretionary funds attached to Research Chairs might be used to underwrite the contract of a junior researcher who is then better positioned to compete for a new investigator award. The chair funds can then be recycled and the process repeated.

CIHR's capacity building efforts in PPH research are not just aimed at increasing the expertise of people already engaged in population and public health research, they are also interested in building capacity for *inter-disciplinary research* by attracting into PPH research researchers from outside the field. Complementing these capacity building efforts on the research side, there is interest also in building *receptor capacity* and in establishing *meaningful linkages* between the research and policy and practice settings. Each of these will require the use of a wider range of metrics than those identified in Table 3 and will also require new metrics to be developed, validated and field-tested.

The Centres for Research Development programme for example is examining how effectively grant recipients have engaged decision makers in PPH research through involvement in the governance structures of each new centre. Recipients of the chairs in applied public health will also be required to document how they have engaged decision makers and tailored their research programs to those decision makers' needs. CIHR's investment in scholarships for students in Masters of Public Health degrees can also be seen as contributing towards a research-literate policy and practice community, and the extent to which it succeeds in doing this (that is, the extent to which MPH trained policy makers and practitioners are more research literate) will need to be examined empirically.

One means of capturing the extent of interdisciplinary work (beyond self-report from grant recipients - the evaluation of the Centres for Research Development Programme asks for information on the number of researchers and trainees new to PPH research for example) is in *co-author analysis* of citations. This is a form of relational or network analysis that documents who is working with whom. The results provide a picture of inter-disciplinary work to the extent that disciplines are reflected in the professional backgrounds of the individual researchers. It does require a special analysis to be completed, however, and underestimates inter-disciplinary work if individual researchers span disciplines.

Relational analyzes of citations beyond co-author analysis can also be used to assess '*research targeting*': that is the extent to which previous research provides a foundation for work to come. *Co-citation analysis* for example indicates who is citing whom and so provides a picture of how new research builds upon what is already known (Moore et al 2006).

4.2.3 Informing Policy

As the case study discussing the ban on tobacco use in public places demonstrated, public policy (and not just the actions of the formal health sector) is important in improving population health. In addition to informing the development of clinical practice guidelines therefore, one would also wish to see evidence of the influence that PPH research has had on the formulation and implementation of healthy public policy. This needs to look beyond the health sector to include policies of other government departments (e.g., employment, recreation, environment and housing) as well as policies implemented by the private sector (e.g., flexible working hours, job latitude, worksite health promotion). This will require occasional mixed-method case studies, document analysis to see whether PPH research is cited, and interviews with policy-makers. There will also be benefit in looking at instances where PPH research evidence is available but is not being implemented, where in fact policy is not evidence based even though it could be, in order to identify cases where the potential return on investment in PPH research is not being fully realized. McLennan & Lavis (2006) for example, have documented how few parenting programs in one Ontario municipality were supported by quality evidence despite this being one area of public health practice that is relatively well endowed with research evidence.

4.2.4 Health and Health Sector Benefits

Much of what counts as PPH research aims to influence risk factors that are common to a large number of chronic diseases. For example, physical inactivity, poor diet and tobacco use are associated with a range of different cancers, cardiovascular disease, diabetes, and in some cases even depression. In addition to the disease–specific health benefits mentioned in Table 3, it will be important therefore to look at *potential years of life lost (PYLL) across multiple disease and 'all-cause' categories*. PYLLs just measure improvements in life expectancy however, and capturing improvements in *quality of life* is also important, especially with the growing burden of chronic diseases and mental health problems. This information is readily available since every major population health survey in Canada since 1990 has recorded quality of life using the Health Utilities Index® (Horsman et al 2003). The Public Health Agency of Canada's Population Health Impact of Canada project has also started to map the relationship between disease burden and quality of life for 200 specific diseases (Flanagan et al 2005).

In addition to the impact PPH research has on health outcomes, it is critically important that we also measure the impact research has on *reducing health inequalities*. Reducing inequalities is integral to the population health approach. It also features as an objective of many of the special strategic initiatives of the CIHR institutes (Aboriginal People's Health, Gender and Health, Health Services and Policy Research, Human Development, Child & Youth Health, Nutrition, Metabolism and Diabetes, and Population and Public Health to name just a few). There will need to be discussion and some agreement reached about what sorts of inequalities should be monitored (are we interested in closing the gap between the richest and poorest groups, or in flattening the gradient across the whole population for example) but the capacity to monitor changes in the pattern of health inequalities already exists (Frohlich et al 2006).

It will, of course, be necessary to spell out what is to be included under *Public Health: Strategic Research Initiatives and their Outcomes* before determining how this will be measured, but by including impacts on quality of life and health inequalities here, as well as capturing the impacts of research on capacity building and policy mentioned elsewhere, we will have covered most of the important outcomes.

4.2.5   Economic Benefits

Recent estimates quantify the burden of disease in Canada to be in excess of $187 billion (Public Health Agency of Canada, forthcoming). Spending on health care comprises $98 billion (or 52%) with the remainder attributed to the loss of economic production as sickness and premature mortality remove people from the labour force. Clearly, even if only a small percentage of this cost can be offset by evidence-informed prevention then the value of population and public health research could be substantial. Several agencies, including Health Canada, and now the Public Health Agency of Canada, quantify health care costs and the impact of disease on productivity and the methods that they use are, to a large extent transferable. The problem of attribution remains however, and in some respects is greater for PPH research than for health services research. The time frames over which both cost-reductions and production effects must be measured are much longer and the scope of the assessment will also need to be broader than just health care costs. As one sees with early child development for example, successful population and public health programmes can affect the costs of education, social welfare and criminal justice as well as health care, and do so progressively over the life time of those affected (Hertzman & Wiens 1996).

The final element in the economic benefits of research is *commercialization*, which refers to both privatization (through the assignment of property rights) and exploitation of the benefits emerging from health research. The effects of commercialization on social value are ambiguous. Privatization of the benefits of health research (as occurs through patents) primarily determines who takes what share of the value that is created: it does not in itself create value. Privatization may increase value, to the extent that it encourages innovation or it may reduce social value by limiting access to the benefits of new technologies in order to maintain a profit maximizing price, as has been seen with antiretroviral drugs for HIV/AIDS in Africa (Chirac et al 2000). Similarly, exploitation of the benefits of research can increase value, (as might be the case when entrepreneurs identify new opportunities to use evidence for social good), or decrease its value as we saw in the discussion on the effect of distortions in health care decision making on the costs of health care.

In any case, commercialization is unlikely to feature highly in any assessment of the social value of PPH research. The impacts of such research often take the form of a public good where the social benefits of research are high but the private benefits are low, and so there is little incentive for commercialization. This is of course, exactly the situation where public funding for research is essential and ought to take priority. If the benefits of research can be commercialized and privatized, then the private sector can (and should) be relied upon to fund it.

4.3 Summary Metrics

In summary, adding to the foundations set by previous CIHR discussions, we have identified a range of measures that can be used to evaluate the outputs and outcomes of PPH research. These include traditional metrics for evaluating the impact of research, such as the number of publications and trainees, and also new indicators, such as the extent to which decision makers have been effectively engaged in the research process of new research centres that are more relevant to the strategic intent of CIHR and other supporters of PPH research. The latter needs to be the subject of further study and resourced accordingly to facilitate the further development and validation of appropriate metrics to measure these complementary impacts of research.

5. Conclusions and Discussion

Canada has had an enviable reputation in population and public health. The Lalonde report, the Ottawa Charter for Health Promotion and the work of scholars associated with the Canadian Institute for Advanced Research (Evans et al 1994) among others have each helped define the field and advance the science. Lack of research capacity and limited research funding have restricted the extent to which we have been able to build on the potential this foundation offered and further orient PPH research towards solutions rather than only documenting patterns of disease and disability and their underlying determinants.

This has now changed. In the eight years since the establishment of the CIHR, it alone has spent perhaps as much as $450 million of public funds to support population and public health research. Canada's real investment in PPH research since 2000 is considerably higher than this as other national bodies such as the Heart and Stroke Foundation and the National Cancer Institute of Canada, as well as several provincial organisations, also fund PPH health research. The total amounts to a substantial investment by the Canadian public, and they can rightfully ask what impact it has had; what return it has generated.

It is imperative that we evaluate the impact of this investment so that Canadians can see whether they are receiving sufficient value. Population and public health researchers cannot escape this scrutiny, but they should not be wary of it. There are two examples featured in this report - tobacco control and the reduction in sudden infant deaths - where PPH research has generated substantial social benefits, valued far in excess of the costs of the original research.

Awareness of the history of population and public health research and its funding in Canada is important in further understanding what benefits we might reasonably expect to see now, and what benefits we can rightfully expect to see in the years ahead. Much of the investment in PPH research by CIHR and its partners since 2000 has gone to build research and receptor capacity. There is still more to be done in this area, but the balance of population and public health research activity is now changing, and it is right that it should do so. The establishment of the Population Health Intervention Research Initiative for Canada (PHIRIC), a multi-agency collaboration to increase the quality and quantity of population health intervention research and its use by policy-makers and practitioners, marks the beginning of a new and complementary emphasis, moving away from research aimed primarily at better understanding of the determinants of health and towards research aimed at making a difference by improving the life chances and health of Canadians.

Of the different methods for evaluating the impact of health research explored here, the payback approach provides the best overall framework. It is comprehensive. It encapsulates the value of health gains that are measured in the return on investment approach, but is capable of being extended beyond this to capture also the impact that CIHR's investment has had on capacity building and policy making. The payback approach is also flexible. It can be adapted to provide frequent, but somewhat shallow appraisals using data that are already routinely collected, and it can be used to provide less frequent, more intensive evaluations as required that involve special analyzes of existing data or new data collected especially for this purpose.

We have also identified several issues. The first is in simply defining the scope of the exercise: what is to count as population and public health research and what needs to be done to validate the assessment of each award so that the investment can be correctly quantified. Since CIHR now commonly partners with other funding agencies it will prove increasingly difficult to limit any review solely to CIHR-funded PPH research.

Secondly, as our case studies amply illustrate, the impacts of health research across the four organizing themes of CIHR will occasionally be synergistic. In extreme cases, there will be no measurable impact unless every component is present. More generally, value will be the consequence of interaction between research themes. In this case, one needs to be especially careful when attributing net returns to particular components of the investment in health research. The logic model associated with the payback approach can help identify where such interfaces occur and so should provide a defence against erroneous attribution.

Third, even where it is possible to correctly attribute health gains to particular types of research (as we did with the health improvements following PPH research and the decline in sudden infant deaths syndrome), it is unlikely that one can easily attribute a proportion of this gain to any one country's investment. Population and public health policy is rarely influenced by the results of a single study. More often it is the weight of evidence that is important, and evidence from several different contexts or countries that all points in the same direction often has more substance.

In conclusion, many existing methods for evaluating the impact of health research tend to be partial in what they evaluate. The best of these partial measures (that is the Return on Investment approach, which is best in the sense that it alone considers the value of the health gains attributable to research) has political value, but the results are unlikely to be universally accepted, not least because of the rather crude way in which the method addresses each of the issues discussed above. In contrast, the payback approach is comprehensive and adaptable, and can be tailored to suit CIHR's many objectives in relation to PPH research beyond knowledge generation. There is a range of indicators readily available that relate to each of the categories included in the payback approach and so a gross picture of the impact of PPH research can be compiled relatively easily. We have also suggested a number of new metrics designed to capture some of the more novel and complex aspects of CIHR funding that can be used in a more intensive exercise to obtain a finer picture of the impact of PPH research.

Some of the new metrics we suggest need development, refinement and validation before such intensive evaluation can occur. In essence, this could be the legitimate focus for a novel program of research. None of the measures are perfect and there are substantial problems associated with the validity and/or interpretation of each one. A more reliable picture may emerge however if the range of measures we identify here are analyzed together. This is a hypothesis in need of testing. It will also be necessary to determine what counts as PPH research so the size of Canada's investment can be quantified.

The biggest challenge remains that of attribution: how much of any measurable impact can be attributed to PPH research and to CIHR's investment within that funding envelope. New research invariably builds on the lessons of past research. Changes in policy and practice are rarely swayed by the results of one study but are influenced instead by the accumulating weight of evidence especially from more than one setting. The impact of research is often synergistic making it difficult to parcel out the separate contributions. None of this should stop us from trying to evaluate the impact of publicly funded research. With the payback approach we can begin to see where research has made an impact and whether this impact was dependent on previous research. We might not be able to come up with precise estimates of return on investment but should be able to show to gross orders of magnitude whether research is contributing more than its costs. Canadians who have pay for the research deserve nothing less.

References

Access Economics, 2003. Exceptional returns: the value of investing in health R&D in Australia. Canberra: Australian Society for Medical Research.

Access Economics, 2008. Exceptional returns: the value of investing in health R&D in Australia II. Canberra: Australian Society for Medical Research.

Buxton, M., Hanney, S., 1996. How can payback from health services research be assessed? *Journal of Health Services Research and Policy*, 1, p. 35-43.

Buxton, M., Hanney, S., 1998. Evaluating the NHS research and development programme: will the programme give value for money? *Journal of the Royal Society of Medicine*, 91 (Suppl 35), p. 2 - 6.

Buxton, M., Hanney, S., & Jones, T., 2004. Estimating the economic value to societies of the impact of health research: a critical review. *Bulletin of the World Health Organization*, 82, p. 733 - 739.

Canadian Institutes of Health Research (CIHR), 2005. Developing a CIHR framework to measure the impact of health research. (Synthesis and Report of Meetings) [Online] Available at: www.cihr-irsc.gc.ca/e/30324.html

Canadian Institutes of Health Research (CIHR), 2008a. 2008-2009 CIHR Grants and Awards Guide [Online] Available at: http://www.cihr-irsc.gc.ca/e/22630.html#1-A1.1

Canadian Institutes of Health Research (CIHR), 2008b. About CIHR: An Overview of CIHR [Online] Available at: http://www.cihr-irsc.gc.ca/e/24418.html

Canadian Institutes of Health Research (CIHR), 2008c. The CIHR Act [Online] Available at: http://www.cihr-irsc.gc.ca/e/22948.html.

Chapman, S., 1997. Tobacco industry memo reveals passive smoking strategy. *BMJ*, 314, p. 1569.

Charlton, B.G., 1995. A critique of Geoffrey Rose's 'population strategy' for preventive medicine. *Journal of the Royal Society of Medicine*, 88, p. 607-610.

Chirac P., et al., 2000. AIDS: patent rights versus patients' rights. *Lancet*, 356, p. 502.

Cutler D., Kadiyala S., 2003. The return to biomedical research: treatment and behavioural effects. In Murphy, K.M., & Topel R.H. (eds). *Measuring the gains from medical research: an economic approach.* Chicago: University of Chicago Press.

Davidson-Rada J., Caldis, S., Tonkin S., 1995. New Zealand's SIDS prevention program and reduction in infant mortality. *Health Education Quarterly*, 22, p. 166-171.

Diener, A., O'Brien, B., Gafni, A., 1998. Health care contingent valuation studies: a review and classification of the literature. *Health Economics*, 7, p. 313-326.

Ducatman, A.M., McLellan, R.K., 2000. Epidemiologic basis for an occupational and environmental policy on environmental tobacco smoke. *Journal of Occupational and Environmental Medicine*, 42 (12), p. 1137-1141.

Dwyer T., et al., 1995. The contribution of changes in the prevalence of prone sleeping position to the decline of sudden infant deaths in Tasmania. *JAMA*, 273, p. 783-789.

Eagan T.M., Hetland J., Aaro L.E., 2006. Decline in respiratory symptoms in service workers five months after a public smoking ban. *Tobacco Control*, 15 (3), p. 242-246.

Eisner M.D., Smith A.K., Blanc P.D., 1998. Bartenders' respiratory health after establishment of smoke-free bars and taverns. *JAMA*, 280, p. 1909-1914.

Evans R.G., Barer M., Marmor T.E. (eds), 1994. Why are some people healthy and others not? New York: Aldine de Gruyter.

Fichtenberg, C.M., Glantz S.A., 2002. Effect of smoke-free workplaces on smoking behaviour: systematic review. *BMJ*, 325, p. 188.

Flanagan W., et al., 2005. Estimating summary measures of health: a structured workbook approach. *Population Health Metrics*, 3, p. 5.

Franks A., et al., 2006. Assessing prevention research impact: a bibliometric analysis. *American Journal of Preventive Medicine*, 30, p. 211-216.

Frohlich K., Potvin L., 2008. The inequality paradox, population health and vulnerable populations. *American Journal of Public Health*, 98, p. 216-221.

Frohlich K., Ross N.C., Richmond C., 2006. Health disparities in Canada today: some evidence and a theoretical framework. *Health Policy*, 79, p. 132-143.

Galan I, et al., 2007. Impact of the "Tobacco control law" on exposure to environmental tobacco smoke in Spain. *BMC Public Health*, 7, p. 224.

Gilman, E.A., et al., 1995. Trends in rates and seasonal distribution of sudden infant deaths in England and Wales, 1988-92. *BMJ*, 310, p. 631-632.

Ginnelly, L., et al., 2005. Using value of information analysis to inform publicly-funded research priorities. *Applied Health Economics & Health Policy*, 4, p. 37-46.

Glantz, S.A., Balbach, E.D., 2000. Tobacco war: inside the California battles. Berkeley, University of California Press.

Goodman, P., et al., 2007. Effects of the Irish smoking ban on respiratory health of bar workers and air quality in Dublin pubs. *American Journal of Respiratory Critical Care Medicine*, 175, p. 840-845.

Hanney, S.R., et al., 2004. Proposed methods for reviewing the outcomes of health research: the impact of funding by the UK's Arthritis Research Campaign. *BMC Health Policy and Systems*, 2, p. 4.

Hanney, S.R., et al., 2006. Identifying the impact of diabetes research. *Diabetic Medicine*, 23, p. 176 - 184.

Hanney, S., et al., 2005. Assessing the benefits of health research: lessons from research into the use of antenatal corticosteroids for the prevention of neonatal respiratory distress syndrome. *Social Science & Medicine,* 60, p. 937 - 947.

Hatfield, M., Sonnenschein, H., Rosenberg, L., 2000. Exceptional returns: the economic value of America's investment in medical research. Lasker Foundation.

Hertzman, C., Wiens, M., 1996. Child development and long-term outcomes: a population health perspective and summary of successful interventions. *Social Science and Medicine*, 43 (7), p. 1083 – 1095.

Horsman, J., et al., 2003. The Health Utilities Index (HIU®): concepts, measurement properties and applications. *Health and Quality of Life Outcomes*, 1, p. 54.

Injury Prevention Committee, Canadian Paediatric Society, 1996. Reducing the risk of sudden infant death. *Paediatr Child Health*, 1, p. 63 - 67.

Institute of Population and Public Health (IPPH), 2004. CIHR – Institute of Population and Public Health Joint Advisory Board Meeting June 10th, 2004. [Online] Available at: http://www.cihr.ca/e/24901.html.

Institute of Population and Public Health (IPPH), 2008a. Mapping and tapping the wellsprings of health: strategic plan 2002-2007. [Online] Available at: http://www.cihr-irsc.gc.ca/e/12274.html.

Institute of Population and Public Health (IPPH), 2008b. IPPH Annual Report 2002-2003. Online Available at: http://www.cihr-irsc.gc.ca/e/23913.html.

Jones-Lee, M.W., 1976. The Value of Life: An Economic Analysis. Chicago, University of Chicago Press.

Kattwinkel J., Brooks J., Myerberg, D. and the American Academy of Pediatrics Task Force on Infant Positioning and SIDS, 1992. Positioning and SIDS. *Pediatrics*, 89, p. 1120 – 1126.

Klose, T., 1999. The contingent valuation method in health care. *Health Policy*, 47, p. 97 – 123.

Kondro, W., 2007. Exit interview with Dr Alan Bernstein. *Canadian Medical Association Journal*, 177(8), p. 844 – 845.

Last J., ed., 2001. Dictionary of Epidemiology. Toronto: Oxford University Press.

Leaders Forum Steering Committee, 2004. Strengthening the foundations of Canada's health research enterprise: a backgrounder. Leaders' Forum for Health Research in Canada, Ottawa.

Lemstra, M., Neudorf, C., Opondo, J., 2008. Implications of a public smoking ban. *Canadian Journal of Public Health*, 99, p. 62 - 65.

Lightwood, J.M., Glantz, S.A., 1997. Short-term economic and health benefits of smoking cessation: Myocardial infarction and stroke. *Circulation*, 96, p. 1089 - 1096.

Lightwood, J., Phibbs, C., Glantz, S., 1999. Short-term health and economic benefits of smoking cessation: Low birth weight. *Pediatrics*, 104, p. 1312 - 1320.

Manuel, D.G., et al., 2006. Revisiting Rose: strategies for reducing coronary heart disease. *BMJ*, 332, p. 659 - 662.

McLennan, J.D., Lavis, J.N., 2006. What is the evidence for parenting interventions in a Canadian community? *Canadian Journal of Public Health*, 97, p. 454 - 458.

Mitchell, E.A., Hutchison, L., Stewart, A.W., 2007. The continuing decline in SIDS mortality. *Archives of Diseases in Childhood*, 92, p. 625 - 626.

Mitchell, V.S., Philipose, N.M., Sanford, J.P., eds., 1993. Committee on the Children's Vaccine Initiative: Planning Alternative Strategies. Institute of Medicine.

Moore, D.S., Haines, V., Hawe, P., Shiell, A. Lost in translation: a genealogy of the "social capital" concept in public health. *Journal of Epidemiology and Community Health*, 60, p. 729 - 734.

Murphy, K.M., Topel, R.H., 2003a. The economic value of medical research. In Murphy, K.M., Topel, R.H. (eds). Measuring the Gains from Medical Research: An Economic Approach. University of Chicago Press, Chicago.

Murphy, K.M., Topel, R.H., 2003b. Diminishing returns: the costs and benefits of improving health. *Perspectives in Biology and Medicine*, 46 (3S), p. S108 - S128.

Organisation for Economic Cooperation and Development, 2007. OECD Health Data 2007: Statistics and Indicators for 30 Countries, OECD, Paris.

Public Health Agency of Canada, 2008. Population health impact of disease in Canada. [Online] Available at: http://www.phac-aspc.gc.ca/phi-isp/summary_measures-eng.php.

Public Health Agency of Canada, 2000. Economic Burden of Illness in Canada. Forthcoming.

Rose, G.A., 1992. The strategy of preventive medicine. New York: Oxford University Press.

Rusen, I.D., et al., 2004. Sudden infant death syndrome in Canada: trends in rates and risk factors, 1985-1998. *Chronic Diseases in Canada*, 25 (1).

Samet, J.M., Burke, T.A., 2001. Turning science into junk: the tobacco industry and passive smoking. *American Journal of Public Health*, 91(11), p. 1742 - 1744.

Schmidt, C.W., 2007. A change in the air: smoking bans gain momentum worldwide. *Environmental Health Perspectives*, 115, p.A412 - A415.

Siegal, M., Albers, A.B., Cheng, D.M., Hamilton, W.L., Biener, L., 2008. Local restaurant smoking regulations and the adolescent smoking initiation process: results of a multi-level, contextual analysis among Massachusetts youth. *Arch Ped Adol Med*, 162, p. 477 - 478.

Siegler, M., Weisfield, A., Cronin, D., 2003. Is medical research cost effective? Response to Murphy and Topel. *Perspectives in Biology and Medicine*, 46 (3S), p. S129 - S137.

Soper, B., Hanney, S., 2007. Lessons from the evaluation of the UK's NHS R&D Implementation Methods Programme. *Implementation Science,* 2 (7).

Stark, M.J., et al., 2007. The impact of clean indoor air exemptions and pre-emption policies on the prevalence of a tobacco-specific lung carcinogen among non-smoking bar and restaurant workers. *American Journal of Public Health*, 97, p. 1457 - 1463.

Strategic Policy Directorate, 2001. The population health template: key elements and actions that define a population health approach. Health Canada [Online] Available at: www.phac-aspc.gc.ca/ph-sp/pdf/discussion-eng.pdf.

Trotter, L., Chapman, S., 2003. "Conclusions about exposure to ETS and health that will be unhelpful to us": how the tobacco industry attempted to delay and discredit the 1997 Australian National Health and Medical Research Council report on passive smoking. *Tobacco Control*, 12 (Suppl 3), p. 102 - 106.

Van Raan, A., 2005. Measurement of central aspects of scientific research: performance, interdisciplinarity, structure. *Measurement,* 3, p. 1 - 19.

Viscusi, W.K., Aldy, J.E., 2003. The value of a statistical life: a critical review of market estimates throughout the world. NBER Working Paper No 9487, NBER, Washington.

Warner, K.E., Hodgson, T.A., Carroll, C.E., 1999. Medical costs of smoking in the United States: estimates, their validity, and their implications. *Tobacco Control*, 8 (3), p. 290 - 300.

Wennergren, G., et al., 1996. The decline in SIDS incidence in the Scandinavian countries and its relation to the impact of risk intervention programs: the Nordic Epidemiological SIDS Study (NESS). Bethesda, MD: Fourth SIDS International Conference, p. 62 - 63.

Wigfield, R.E., et al., 1992. Can the fall in Avon's infant death rate be explained by changes in sleeping position? *BMJ,* 304, p. 282 – 283.

Wooding, S., et al., 2005. Payback arising from research funding: evaluation of the Arthritis Research Campaign. *Rheumatology,* 44, p. 1145 - 1156.

Zitt, M., 2005. Facing diversity of science: a challenge for bibliometric indicators. *Measurement*, 3, p. 38-49.

Meso-Level Metrics for Impact

Metrics for the Treatment Sector or Meso Level of the Canadian Health Care System

Jerald Hage, Director
Center for Innovation
University of Maryland
College Park, MD 20742
United States of America

Preface

The objective of this paper is to propose a metric system for evaluating the return on investments (hereafter ROI) from medical research in the Canadian health care system. The strategic choices of specific metrics in the system are discussed so evaluators can understand the logic that undergirds the health care system. This paper will indicate how one can measure both the impact of a discovery and the impact of its diffusion.

Executive Summary

The sub-title of the document Prospectus for a Major Assessment: The Return on Investments in Health Research: Defining the Best Metrics (Canadian Academy of Health Sciences 2007) raises an implicitly, if not explicitly, important question: what are the best metrics for measuring ROI? This in turn leads to another fundamental question: by what criteria should one evaluate metrics that are used to measure ROI?

To answer these two questions, I suggest that evaluation metrics for medical research (and beyond this, evaluation metrics for ROI in scientific and technological research) should have a number of important characteristics. They should be simple to use and yet fine-grained enough to capture small improvements in health care resulting from specific research studies. At the same time, they should provide the capacity to assess total investment in medical research, i.e. they should be able to function at both microscopic and macroscopic levels. Since there are a number of stakeholders within the Canadian context, the range of metrics needs to include not only economic benefits but societal ones as well, a point well recognized in the synthesis report of the meetings held in Ottawa, Canada to discuss the development of a metric system (Canadian Institutes of Health Research 2005). This system of metrics should provide a considerable amount of feedback to policy makers and, in particular, point to how obstacles and blockages in the system can be eliminated. Too often evaluation studies do not stipulate how ROI can be increased, which, given increasing competition for research funds between

different and vital sectors of science and technology, is a desired outcome. Finally, this metric system should reflect the latest advances in evaluation research so that it represents the 'state of the art'.

Section One: Introduction: Strategic Choices and an Overview of the Metric System

The Advantages of Starting with the Strategic Choice of a Treatment Sector or Meso Sector Level for an Evaluation of the Health Care System

The first and most critical choice an evaluator must make is the level of analysis at which an assessment will be carried out at. Many systems that evaluate scientific research and industrial innovation are constructed at the macro level. Indeed, the focus of the document Developing a CIHR Framework to Measure the Impact of Health Research (Canadian Institutes of Health Research 2005) and the logic models within it are largely at this level (see in particular p. 27). The macro level is critical for both practical and theoretical reasons. On the practical side, this is the level at which policy makers' shape and debate policies. On the theoretical side, the macro level of evaluation fits within an institutional framework called the National Systems of Innovation (Nelson 1993).

While recognizing the importance of these characteristics of the macro level, my proposed system of metrics begins with the treatment sector level within the health care system (see Table One for definitions of these key concepts) and then suggests how data collected at this level can be aggregated to the national or macro level of policy making, the initial focus of Developing a CIHR Framework to Measure the Impact of Health Research (Canadian Institutes of Health Research 2005). The focus on the treatment sector level retains the advantages of the logic models described in the Canadian Institutes of Health Research (CIHR) document. At the same time, the focus on the treatment sector level provides more information that can be used by policy makers. Within health care, the most appropriate term for this level is the **morbidity sector**. Morbidity represents the incidence of a particular disease, injury or health condition.

Because the greatest variation in morbidity occurs at the treatment sector level, it is the level at which data should be collected. But what is the treatment sector or meso sector level? The meso sector level is defined by the *differences* in treatments: for genetic defects (e.g. alpha minus one deficiency), injuries (e.g. post-traumatic stress syndrome), illnesses (e.g. breast cancer), and degenerative processes (e.g. Alzheimer's disease). An important assumption is that medical research varies by the nature of a health care problem and its implied treatment. 'Treatment' is used here in the broadest sense of the term and includes prevention and biomedical and population research that expands knowledge that can be essential for developing treatment strategies.

Since the greatest variation in the nature of medical research occurs for specific kinds of treatments, this becomes the most effective level that can provide feedback to policy makers, who need specific policies to effectively make decisions rather than one general policy that might exclude some areas of medical research and their respective treatment sectors. Focusing on the meso level and aggregating to the macro level provides temporal flexibility to policy makers and evaluators: it can handle both short- and long-term assessments. Additionally, it provides flexibility for the focus of the evaluation. Since the treatment sector is also defined by the similarity of treatments or technologies and the homogeneity of clients or customers, one can select different levels of homogeneity for evaluating ROI from medical research. One can make finer and finer distinctions between the similarity of treatments and patients. One might choose to focus on bullus emphysema at one level, one of its causes, alpha minus one

deficiency, a sub-division of the causes and thus an increase in the homogeneity of the patient pool. This distinction becomes important when the research is concentrating on only one of the potential causes while developing an effective treatment, as is frequently the case. The meso sector level of metrics also allows policy makers to focus on only one stream of research and evaluate its benefits - scientific, economic and societal - if so desired.

But the greatest advantage of analysis at the meso level is the simplicity it provides in making attributions that connect particular research studies and their subsequent impacts on health care (and beyond it, the various economic benefits that accrue). It becomes much easier to make attributions or linkages between investments in a particular kind of research and its potential, or actual pay-off, for both health benefits (better diagnoses), economic benefits (reductions in treatment costs) or societal benefits (reduction in working days lost) (Canadian Institutes of Health Research 2005, p. 2-3). This document (p. 8) expresses concerns about tracing the linkages between research outputs and health care impacts, especially if knowledge develops incrementally over an extended time period. Correct attribution cannot be made at the macro level, where, even within the same disease category, various research studies are confounded. Focusing on the treatment sector or meso level of the health care system solves this problem.

Evaluations of scientific research (including medical research) are caught on the horns of a major dilemma. On the one hand, an increasing desire for accountability necessitates quick assessments of pay-off to provide feedback to policy makers. On the other hand, the desire for an expanded assessment of societal benefits of scientific research requires a longer temporal horizon. To solve this dilemma, we must recognize the difference between *potential* pay-offs from medical research and *actual* benefits (which can only occur when new knowledge is widely diffused throughout the health care delivery system). In evaluation research, this is recognized as the distinction between a research discovery or finding, its dissemination, and its complete diffusion within the health care system. The distinction between measuring the impact of a discovery (advances in scientific knowledge) and the impact of its diffusion (organizational learning) is discussed in Section Two: Metrics of Health Care Impact.

Collecting data at the treatment sector level does not mean that various macro level indicators of performance (such as contributions to knowledge or broader economic benefits) are lost to view - quite the contrary. These are aggregated across the different treatment sector levels, providing flexibility to not only policy makers but also evaluators. Strategically, the treatment sector level lies between the macro level of policy makers and the micro level where research is conducted and patients are treated. It is the missing link that connects these two levels and without research at this level, one cannot easily select the correct health care policies or understand what the various kinds of blockages and obstacles to the creation of new knowledge might be. The most important strategic advantage of developing health, economic, and societal indicators at the treatment sector level is that this ensures simplicity when coding the results of specific research studies.

The treatment sector level comprises the treatment process. It can be broken down into four stages: prevention; intake and assessment (diagnosis and prognosis); treatment; and post-treatment. Medical research typically impacts on only one of these stages and seldom on the major outcomes desired by stakeholders (such as increases in the average duration of life given the morbidity (QALYs)). Thus, viewing the treatment as a process allows one to observe the many incremental steps towards these final objectives. Therefore, this analytical level provides a fine-grained system of health care metrics that parallels the treatment process, as is indicated in Section Two.

Recognizing the differences between treatment sectors allows policy makers to better discern which areas of investment in medical research are more likely to have the highest pay-off (measured in terms of various economic and societal benefits) *provided that they can estimate the amount of time and effort needed to achieve a particular research output*. This is called **prospective evaluation**. How this might be accomplished is discussed in Sections Two and Three. With this information, policy makers can plan more effectively (Canadian Academy of Health Sciences 2007, p. 5).

This tight link between medical research and health care impacts in the treatment process is not the only advantage of analyzing at a meso level. Another is that it calls attention to how medical research for a specific treatment is organized. Borrowing ideas from the well-developed literature on how knowledge evolves in scientific-industrial innovation literature (Hage & Hollingsworth 2000; van Waarden & Oosterwijk 2006) provides insights into the processes of differentiation in medical research organizations, in particular dedicated institutes, which may lead to gaps that are slowing down the development of radically new treatments. If so, this becomes a critical kind of feedback to policy makers.

Definitions of the different parts of the health care system are listed in Table One, indicating how complex this system is. The different meso treatment sectors can be aggregated into the macro or national system of health care. The different meso research sectors (one for each of the major treatment sectors) can be aggregated into the macro or national system of medical research. The same could be done for training programs. More critically, although the main focus in this white paper is on the stages of the treatment process, the stages of the medical research process are indicated as well because this impacts on the speed with which new treatment protocols are developed and more knowledge about the functioning of the body is accumulated. The former issues are discussed in Section Two while the latter are the focus of Section Three.

The distinction between potential and actual benefits provides a solution to help meet several intellectual objectives of CIHR (2005, p. 2). Potential impact of research findings measures advances in medical knowledge. Actual impact of research findings measures the amount of capacity present in the system. As indicated above, when selecting a metric system, an important criterion to consider is the number of intellectual problems the system could solve.

This connects to the final criterion for the evaluation of a system of metrics based on the differences between treatment sectors: whether it represents the 'state of the art'. Two new thrusts in evaluation have been advocated by European researchers (Arnold 2004). The first thrust is the importance of identifying blockages and obstacles that prevent the rapid development of new innovations. As indicated in Section Three of this white paper, these blockages and obstacles are the probable causes of the gaps evident in the rapid development of new treatments. Although Arnold's focus is on the evaluation of science and technology, the same logic can be applied to how treatments are organized and whether or not the different stages in the treatment process are well connected. Some of the metrics suggested in Table Two in Section Two indicate how this might be assessed.

The second thrust is the importance of allowing theory to inform evaluation. Hopefully evaluations of science and technology (including medical research) will start contributing to the construction of theory. While evaluations of medical research have their own specificities, one reason this white paper stresses the importance of studying at the treatment sector or meso sector level, it is also true that different

theoretical formulations such as the Idea Innovation Network Theory (Hage & Hollingsworth 2000) can provide cognitive maps that raise important questions and provide new insights. These are the advantages of this proposed system of meso level metrics for determining ROI from medical research. In other words, new developments in social science theory can inform evaluation and evaluation can help contribute to the development of new social science theory. We already have some suggestions of this in the ways in which advances in knowledge and organizational learning can be assessed with this system of metrics.

Table One: Definitions of Key Concepts

Key Concepts	Definitions
Macro level of analysis of health care	The aggregation of different treatment sectors, research programs, and training programs, and health care policy makers and decision makers
Meso level of analysis for treatment	Variations in treatment processes defined by differences in technology, procedures, and target populations
Stages in the treatment process	Prevention, diagnosis and prognosis, treatment, and post-treatment including long-term care, and knowledge about the functioning of the body
Meso level of analysis for research	Variations in research programs defined by the differences in the health care impact and/or target population
Stages in the research process	Basic research, clinical research, protocol development, research on service provision and quality of patient care, research on the dissemination and diffusion of research results
Micro level of analysis for treatment	Hospitals, clinics, rehabilitation centres, and other facilities that provide patient care of one kind or another
Micro level of analysis for research	Universities, medical schools, research institutes, research hospitals, industrial research laboratories and other research units involved in health care research

Overview of the Metric System

As noted in the executive summary, medical research is multi-faceted and should be evaluated across different dimensions. The specific metrics included in this white paper tap into these different dimensions:

1. Metrics of health care impact by stage in the treatment process;
2. Metrics of research investment by arenas within the production of medical knowledge within the specific treatment sector;
3. Metrics of contributions to scientific knowledge;
4. Metrics of network gaps in the production of innovative treatment protocols;
5. Metrics of economic and social benefits of medical research.

The issue is how these different dimensions relate to each other. The discussion of these metrics does not follow processes provided by the CIHR logic models (2005, p. 27) but this is for a specific reason. A typical logic model would probably begin with the metrics of research investment (as these provide the resources for the creation of health care impacts) and then move to the metrics of health care impacts. I have reversed this logic because the major issue is: what are the indicators of health care impacts? (And in particular, what are the advantages of a fined-grained approach to the treatment or meso sector level of the health care system?) Demonstrating the ROI of medical research is the most important part of this exercise and it depends upon the variety and number of health care impact indicators one has developed.

Throughout the discussion of these metrics, a continual concern is to provide a number of policy feedbacks so that the performance of the health care system can be improved and ROI from medical research can be increased. Indeed, as I have indicated, one of the advantages of studying the treatment sector or meso level of the health care system is the emergence of useful feedback to policy makers. This includes the identification of whether or not some components of the treatment process are ignored, the arenas of medical research that receive little investment, and the gaps in the production of radical new treatments in the networks connecting different kinds of medical research.

However, it is important to stress that I do not provide any metrics for measuring whether research findings have changed policy in any way. I perceive this to be a different set of issues and well discussed in Hanney (2007) and Borbey (2007). *Here the focus is on what the feedback should be rather than whether the feedback changed policy.* The real issue is to provide meaningful feedback to policy makers. Examining different treatment sectors by focusing on the meso level rather than the macro level is much more likely to provide this kind of feedback. As indicated above, it is important that the feedback contain not only information about ROI, but more critically, data that would inform changes in policy and in particular, reduce blockages and obstacles that slow the development of quick and effective treatments.

The elaborateness of this list of metrics speaks to the varied concerns of different stakeholders, while simplifying the categories of research 'pay-back' proposed by CIHR (2005, p. 1), which are consistent with those of Buxton and colleagues (1994) and as modified by the Canadian Academy of Health Sciences (n.d.): Advancing Knowledge, Informing Decision-Making, Health Impacts, and Economic Impacts.

What this white paper does not accomplish is also important to state. The paper is focused on developing metrics, not the methods that one would use to implement an evaluation using these metrics (see Buxton 2007; Hanney 2007; Wooding 2007 for various examples of methods). This would require another white paper, particularly as there are a number of alternative research designs that could be used to collect the necessary data. In addition, CIHR (2005, p. 32) has an extended list (see Hanney et al 2004). At various points in this paper, suggestions about methods are made but it is not considered to be part of what this white paper is intended to accomplish. Nor does this white paper consider the issue of how to evaluate relative priorities of medical research in comparison to research in the physical sciences, the military or any other national goals of the Canadian government. However, the logic of the process used in this white paper can be applied to other national goals - their delivery systems and research investments. If this were done, then one could compare the relative ROIs in different national sectors of concern to the government. Clearly, this is considerably beyond the scope of this present exercise.

Section Two: Metrics of Health Care Impact

Economists have developed an elaborate classification system of industrial product sectors and these sectors include market niches. But this effort has not yet been applied to non-economic service sectors such as health, education, and welfare, to say little about new national missions such as national security, global warming, etc. Despite this, the same logic for distinguishing between product markets can be applied to the classification of service sectors.

It is the difference between the kinds of patients and treatments (including technologies such as machines, procedures and human expertise) that allows one to observe distinct treatment or morbidity sectors. Some might question the comparison between the treatment sectors in the health care system and the industrial sectors within the economy. But by thinking in terms of analogues, one can develop a number of insights. Admittedly, the insights must always be carefully adapted to the specific circumstances to observe some of the more striking differences as well.

One might ask, why distinguish separate sectors within the health care system? Since many different treatments are housed in the hospital, the concept of distinct treatment or morbidity sectors may appear to be strange to health care professionals. But it is precisely because of this fact that one needs to clearly define treatment sectors so that one can more easily establish a linkage between a body of research or specific finding and its health care impact, even if much of the diagnostic equipment resides in the same place as other treatment sectors. Indeed, this is one of the more interesting planning issues, especially given the problem of intake, when it may be better to provide specialized clinics for particular stages in the treatment sector. Recognizing the alternative treatment systems allows for comparisons across the research findings of the thirteen Canadian Institutes of Health Research, some of which reflect particular kinds of populations (Aboriginal, elderly, youth), some of which reflect specific systems within the body (circulatory, neurological, musculoskeletal) and other specific arenas of research (basic such as genetics or service provision as in the Health Services and Policy Institute). Despite the name of a specific institute, it might be involved in research relevant to other institutes.

But this is not the only reason to focus on the meso sector. Highlighting the treatment sector or a specific morbidity calls attention to parts of the treatment process that may require strengthening, for

example, via research on service delivery in the Institute of Health Services and Policy Research. At minimum, four components or stages in the treatment process can be discerned: the prevention stage; the intake and assessment stage (including diagnosis and prognosis); the treatment stage (including hospitalization); and the post-treatment stage (including rehabilitation and long term care when appropriate). Prevention is placed prior in time with the simple assumption that if prevention can be successful, then the treatment process is unnecessary.

Metrics for the Four Stages: Prevention, Intake and Assessment, Treatment and Post-Treatment

(Including Long Term Care)

Carefully specifying the stages in the treatment process associated with a particular morbidity allows for a fine-grained set of health care impact metrics or indicators. One could make additional distinctions within these four stages. For example, one might want to distinguish between diagnosis and prognosis. Improving the quality of a prognosis allows individuals to decide that continued treatment is not necessarily worth the effort, especially if it degrades the quality of life. Government policy makers might decide to ration certain interventions given the prognosis for a specific age group (e.g. as has the U.K. - no kidney transplants for individuals over the age of 45; the tendency in the U.S. to refuse to perform prostrate surgery in men over the age of 70 because of their life expectancy). On the other end of the continuum there is the question of the appropriateness of treatment with mild severity (Canadian Academy of Health Sciences 2007, p. 12). Given the soaring costs of health care and acknowledging that policy makers do not like to use the 'r' word – rationing - the reality exists that a strict cost-benefit analysis might exclude certain interventions in particular age groups.

The four stages used to describe the treatment process are the same as those of CIHR (2005). Each of these four stages suggests metrics of health care impacts from research (as indicated in Table Two). One must start with a fine-grained conceptualization of the treatment process to be sure to capture the specific impacts of particular research findings. Although the methods used to make assessments are not part of this white paper, the intent of discussing these metrics is for an evaluator to read research findings in a project report and code them in terms of treatment impacts. Furthermore, the impact must be weighted in those instances where the gains are limited to a certain percentage of the patients, which is quite typical in most treatment interventions.

In addition to the four stages of the treatment process, I have added a category called 'Knowledge About the Health Care Problem', because a major part of biomedical and population research focuses on the development of understanding a health care problem that eventually can lead to either prevention or treatment. The suggested three metrics for measuring impact of health care research are discussed below as a special issue. Other important measures of knowledge are suggested below and in particular in Section Three. Another important, special issue for some health care professionals is the impact of research on quality of life. Indicators for this have been added. Finally, research on health service delivery not only potentially impacts on quality of patient care but also on the speed with which diagnosis and treatment occur. Again, several indicators that reflect this factor have been added.

The treatment process logically begins with the stage of Prevention. In this stage, I suggest two metrics for measuring health care impact. The first and most obvious one is the relative effectiveness of

prevention. Prevention protocols that stopped smoking and discouraged teenagers from starting smoking have steadily improved over the years. As more people adopt healthier eating and exercise habits, one observes the decline in the severity of various health problems associated with aging. Perhaps the most dramatic examples of prevention are the development of vaccines that eliminate specific morbidities such as small pox, polio, and, we hope, some day AIDS.

Intake and Assessment is a particularly interesting stage. Within this category, I suggest three metrics for measuring health impact. The speed of diagnosis can be strongly impacted upon by either the adoption of highly specific screening techniques and/or the rearrangement of the delivery system of health care so that there are more points of contact at which a quick diagnosis can be made.

Table Two: Metrics of Health Care Impact

Stage of the Treatment Process	Metric
Prevention	**Percent increase in the effectiveness of prevention intervention (decline in incidence of morbidity)**
	Percent decline in severity of incidence of morbidity
Intake and Assessment	**Percent increase in the speed of diagnosis (reduction in the number of tests and their duration)**
	Percent increase in the accuracy of diagnosis (reduction in false positives or negatives)
	Percent increase in accuracy of prognosis (duration and quality of life, etc.)
Treatment Interventions	**Percent increase in the speed of treatment intervention (wait time in the emergency room)**
	Percent decrease in the length of treatment
	Percent decrease in side-effects of intervention and/or their severity
	Percent decrease in opportunistic infections during treatment intervention
	Percent increase in the quality of life during treatment (reduction in invasive procedures, increased opportunities to be treated as an outpatient, reduction in pain during major interventions, etc.)
	Percent increase in success rate of intervention
Post-Treatment Interventions (Rehabilitation and Long Term Care)	**Percent increase in the speed of the rehabilitation intervention**
	Percent decrease in length of rehabilitation and long term care
	Percent increase in the quality of life during rehabilitation and after care (decrease in the pain of rehabilitation procedures, reduction in invasive procedures, opportunities to be treated as an outpatient)
	Percent increase in success rate of rehabilitation (DALYs) or increase in physical (vision, hearing, thinking, movement, dexterity) and psychological functioning (cognitive processing, speech, memory) after stroke or injury that impaired functioning
Summary Output Measures of the Morbidity Sector	**Percent increase in the average duration of life given the morbidity (QALYs)**
	Percent increase in the quality of life after interventions (reduction in recurrences, continuity in mobility, reduction in constraints of life style, etc.)
Knowledge about the Health Care Problem	**Percent change in the understanding of the causes of the health care problem**
	Percent increase in the sub-categories of the health care problem
	Percent increase in the understanding of the relevant biological and psychological processes of the body relevant to the health care problem

The importance of eliminating false negatives and false positives in diagnosis is well understood, as is the importance of the accuracy of a prognosis. Accurate prognoses can lead to quite different decisions about whether it is worth starting a treatment or not. Many patients will accept a death sentence provided there is a good control of pain. Much research has cast doubts on the advantages of screening precisely because of inaccuracies of the diagnoses - too many false positives and false negatives (e.g. Marshall 2008 on mass screening of high risk lung cancer patients). But gradually through decision analysis (such as Goldie's work), considerable progress is being made in how to effectively screen for certain cancers.

Again, I begin the third stage, Treatment, with a measure of the speed with which an intervention occurs. As is well known, after a heart attack, an intervention within two hours considerably increases the likelihood of recovery. But what determines an intervention within two hours? Research on the nature of the delivery system can impact decisions about the availability of access to a cardiac unit and what a patient's best placement is. Percent increase in the success rate of an intervention is an obvious over-all measure of this stage. Equally important is the reduction in side effects as a consequence of an intervention and elimination of opportunistic infections, a growing problem in American hospitals. Then, too, reduction of time in treatment has an obvious impact not only on cost but also on the perception of the quality of life.

The same logic for Treatment applies to Post-Treatment. The sooner that some form of rehabilitation (where it is relevant) begins, the more likely it is that a successful outcome will occur. Reductions in the length of the rehabilitation period and in the amount of long-term care can have enormous impacts on the cost of health care, especially as the Canadian population ages with increasing frequency of stokes, Alzheimer's, and other degenerative conditions.

One can use two summary measures of the four stages: the average increase in the age-sex standardized duration of life for a specific morbidity and an increase in the quality of life after intervention. The former measure is usually only useful when evaluating a stream of research because most research studies do not impact on this - and if they do, it is incrementally, if at all. The second summary measure, the quality of life after intervention, is discussed below in a special sub-section.

For those who want to demonstrate the value of medical research in real time, the focus must be on some of the measures in the treatment stage. Emphasizing the metric 'duration of treatment' encourages research that reduces this, such as laparoscopic surgery for gall bladders, yet this procedure has little effect on the duration of life. However, it does considerably reduce costs and increase quality of life, other metrics listed in Table Two. And this is the kind of feedback that policy makers need. Where are the best leverage points in research funding? Are they ones that can impact on several of these health care metrics and beyond this to, of course, reduce overall health care costs within a specific morbidity?

Research projects in the disparate institutes can concentrate on one or another stage of the treatment process. The distinction by stages allows us to recognize their contributions to the overall health of the

Canadian population. For example, the improvement in the duration of life of individuals with cardiovascular problems is a consequence of three distinct kinds of changes: faster and better medical interventions after a heart attack, drugs to reduce some of the causes of cardiovascular problems such as cholesterol and blood pressure, and behaviour changes involving diet and/or exercise (Canadian Institutes of Health Research 2005, p. 13). Each of these different contributions is captured in this system of health care impact metrics.

However, what is usually *not* demonstrated in these kinds of evaluations is an indication of which countries have made what contributions and how important particular national streams of research are. There are presently software packages that can trace back the citations that are most important in the development of these advances in medical knowledge. Also in computing the cost-benefit analysis, usually the cost of medical research on this problem in all the countries that are involved (and especially those that are responsible for some of the more important citations) are not added together, which would dramatically change the cost-benefit ratio.

At the meso sector level, these metrics are highly flexible and can be used with micro time, say three to five years or with macro time, three to five decades. In the former instances, one would examine a few research studies and in the latter, one can evaluate how an entire program of research, e.g. cardiovascular disease treatment, has unfolded over three or more decades (Canadian Institutes of Health Research 2005, p. 10-11). As was suggested at the beginning of this white paper, policy makers in certain circumstances would like immediate feedback relative to options as to how money should be invested in health research. This fined-grained approach allows for this.

At the same time, health care professionals may want to use an assessment of the contributions to duration of life of research that extends over a period of three or four decades for public relations purposes. But my recommendation is that to give proper credit to the various contributions that have been made by researchers in different countries, one should follow the pattern of citations back into time. This would clearly indicate the particular points at which Canadian research added to the pool of knowledge and how this helped others make additional contributions. The Department of Energy in the United States has developed a software package that can trace important citations of patents backwards in time that could facilitate the assessment of the contributions of individual countries. In addition, the fined-grained approach demonstrated in Table Two allows for the detection of contributions that might have been missed.

In practice, one might not want to use this much detail in an accounting scheme, but it is better to start with details and then collapse metrics (and even stages) afterwards so that the coding of research findings becomes easier than it would be if one only focused on the overall metrics listed in Table Two, which is the more typical pattern (Canadian Institutes of Health Research 2005, p. 10-11). As various studies of medical research have indicated, large impacts on QALYs are relatively rare. And while cancer research in general has not resulted in a large reduction in mortality rates, significant progress has been made in some kinds of cancers, again indicating why it is important to study the specific treatment sector and select various degrees of homogeneity, as I have already discussed. Finally, one advantage of this detailed approach is that it allows evaluators to focus on a few indicators that are more sensitive to

the coding of research studies than QALYs, an issue to which I return to in Section Four where economic and societal benefits are discussed.

Special Issue: Knowledge about the Health Care Problem

Technically speaking, medical research that increases knowledge about a health care problem does not necessarily have an immediate impact on the various stages of the treatment process. It is, however, a critical component of medical research and can influence the direction of clinical research. Therefore, this is treated as a special issue in measuring the impact of medical research on health. Although more categories of knowledge can be added, I propose starting with three: increased information about the etiology, recognition of additional types or sub-categories of the health problem, and greater understanding of the physical and psychological processes in the body that are relevant to a specific morbidity. Both biomedical research and population research can cast light on the etiology of a particular health problem. How avian flu becomes a human flu is a current example of the former, while population research on various kinds of destructive behaviours such as drug addition, eating disorders, reckless driving, and unsafe sex are illustrations of the latter.

Even research that focuses directly on the causes of a health care problem typically finds over time that the initial understanding of it is too simplistic. Research on the genetic causes of cancer is a good example. Researchers have gradually indentified different kinds of genes with different functions, including genes that affect the immune response as separate from genes associated with the appearance of cancer. This reflects a movement towards the recognition of greater complexity in understanding the causes of a health care problem.

Progress in research about a health care problem also is the recognition of different sub-types within a health problem, e.g. the movement from hepatitis A to B and C. Or more recently, treating alcoholism with drugs now requires different therapies for different individuals (Miller 2008). Again, this movement echoes another general pattern in the evolution of medical research, namely the recognition of the need for the customization of treatments.

Perhaps the greatest demonstration of these two evolutionary processes towards more complexity and the need for more customization in treatments is reflected in the stream of research about the biological functioning of the body, which provides the background for doing clinical research and may, in the long term, lead to more effective interventions. The clearest example of this is the double helix description of DNA and the demonstration of RNA followed by the decoding of the genome, which in turn has led to a renewed recognition of the complexities involved. Molecular biology has opened the door to the development of gene therapies and the movement towards customization of treatments. Some recent examples of the kinds of general research in the biology of the body reported in Science recently included research on the self-organization of proteins (Lutkenhaus 2008) and the enigmas of blood clot elasticity (Weisel 2008).

Special Issue: Measuring the Quality of Life

Some health care professionals would like to have a global measure akin to health status that indicates the contribution of medical research to the perceived quality of life. I am assuming that it is understood that this is the perception of quality of life as a consequence of health status and not as a consequence of either income status or the nature of important social relationships. Declines in income or the loss of loved ones have strong impacts on the perception of quality of life. The same is true for loss of health.

Rather than use some form of an attitudinal survey, which of course is one way of measuring both perceptions and attributions (whether health, income or relational), I prefer to advocate measurement by relatively firm behavioural measures that are combined into an overall measurement of an objective improvement in quality of life. The second procedure for measuring quality of life is to deconstruct this large global measure into three separate measures: during treatment, during post-treatment including long term care, and finally after the completion of health care interventions of any kind relative to the same morbidity.

What are the behavioural indicators that can successfully model the perception of a high quality of life? Among others for measuring quality of life during treatment are reductions in invasive procedures, opportunities to be treated as an outpatient, reduction in pain during and after major interventions, etc. One example of the first indicator is the substitution of laser surgery for the eyes rather than more invasive forms of surgery. The use of drugs to prevent surgery has recently been discussed in the case of prostate cancer (Kolata 2008). But in this instance the potential long term side effects are unknown and unlikely to be studied. The movement of treatment out of the hospital into the outpatient clinic also has the same consequence. Most people find hospitals to be scary places, to say nothing about their inherent dangers associated with opportunistic infections. Finally, the successful management of pain is a critical issue most patients have. When pain is reduced without side effects such as dependency, most patients would consider this as an improvement in quality of life. Speed of treatment and post-treatment rehabilitation probably also can contribute to the perception of improved quality of life since both waiting (including waiting for the diagnosis and prognosis) are stressful times for the patient and his/her family.

Behaviour indicators of quality of life once an intervention is finished are somewhat different, yet analogous and include reduction in recurrences, continuity in mobility (broadly defined to include all functions), reduction in constraints of life style, etc. Simply put, being able to return to one's previous patterns of life is the best indication of quality. As is well known, many morbidities necessitate considerable constraints on life style if patients are to avoid recurrences. Thus, patients with destructive habits cannot experience a high quality of life unless ways are found to protect them from themselves.

In each of the stages, the metrics follow clear patterns. Both speed and quality of life are two common metrics. I have suggested metrics that tap into the quantity of improvement and metrics that measure the quality of improvement. The category of quantity improvement includes the metrics of effectiveness of prevention, accuracy of diagnosis and prognosis, success rate of intervention and success rate of rehabilitation whenever it is relevant. The latter category includes decline in severity of incidence, decline in side-effects and opportunistic infections during treatment, length of time for treatment, rehabilitation and long-term care (for an argument as to why one should have different kinds of metrics when developing an index, see Hage 1972). One could add more metrics but twenty would appear to provide a fine-grain design to adequately measure contributions from research.

 A research project might have implications for more than one aspect of the treatment process, e.g. a new experimental treatment for melanoma at the National Institute of Cancer in the United States

considerably reduces the amount of time spent in hospital, increases the success rate of treatment from 15 percent to 50 percent, and improves the quality of life during the treatment. Under these circumstances, one must add the percent change in each metric to capture the complete treatment impact. Multiple impacts would be indicative of a major breakthrough in treatment. However, it is also important to recognize that, in this case, the breakthrough occurred after some twenty years of continued research by Dr. Rosenberg and his teams in which there were many dead-ends and continued learning. The sudden leap in progress would probably not have been possible without this prior effort, again raising questions as to how best select the time period for evaluating the ROI, an issue raised by Buxton (2007). My recommendation is to evaluate a specific research study, but at the same time recognize prior effort and research in the assessment.

As in the economic classification of industrial sectors, the number of treatment sectors one is able to distinguish is the issue. From a planning perspective, pragmatism is desirable. Certainly the research interests of the various research institutes in Canada and the priorities of the Canadian Institutes of Health Research represent a useful starting point. As can be seen from this list, the advantage of focusing on treatment or morbidity sectors is that this allows for considerable flexibility in comparing and contrasting research findings from different research programs in the distinct Canadian institutes. One reason for adding indicators of speed and knowledge about a health care problem is to capture the efforts of biomedical researchers, population researchers, and researchers that study the provision of health services. On the other hand, one could sample only a few streams of research, for example, ones that were of particular interest to either policy makers or the public.

Potential versus Actual Benefits

The distinction between potential and actual benefit speaks to two quite different policy objectives in measuring ROI. The potential impact is a measure of the amount of scientific learning or the extent of knowledge advance that has occurred from a specific research study or even an entire research program. The actual impact is a measure not only of the extent of diffusion of knowledge but one which also reflects capacity building in the health care system. The more health care personnel that have learned new protocols, the greater the advance in human capital for that specific morbidity treatment system. Another advantage of clearly separating these two kinds of indications of 'pay-back' is that they address some interesting problems in the sociology of science and organizational literatures. *Advances in treatment knowledge represent a measure of scientific learning while diffusion of treatment knowledge represents a measure of organizational learning.* Both of these measures are different from those listed in Table Two, all of which focus on the knowledge background of a specific morbidity.

Evaluating impacts of medical research on the treatment or morbidity sector must confront the distinction between potential impact and actual impact within the health care system. For example, a research finding on recurrent melanoma may indicate that life can be prolonged three months in about 20 percent of the cases. This reflects the potential benefit if the research finding is diffused throughout the health care system. But this potential benefit can only be realized if all oncologists learn about this new research finding and more critically learn the intricacies of the treatment protocol, which sometimes can be quite complicated. To achieve the actual benefit necessitates measuring how far a specific research finding has diffused throughout the health care system. For this reason, in the next section the research arena labelled 'commercialization' is considered a critical arena because it focuses on how best to diffuse advances in medical knowledge so that potential benefits are actually realized.

Probably the major stage in which the difference between potential and actual benefit is the greatest is the prevention stage. In many morbidities, considerable information exists as to how to prevent illnesses such as AIDS, lung cancer or diabetes and certain accidents (drunken driving), but realization of the actual benefits necessitates enormous changes in human behaviour, most of which are unrealistic. But to call attention to this aspect of the health care system, I have added the metric 'the effectiveness of prevention' to highlight the importance of research that finds intervention methods that can change behaviour, eliminate genetic disorders, prevent epidemics or slow the degenerative processes of aging.

My recommendation to the Canadian Academy of Health Sciences is to measure both potential and actual impacts. The obvious advantage of measuring both kinds of impacts is that the latter allows one to assess whether diffusion of knowledge is incomplete and if so, to begin to identify reasons for this. This, of course, is clearly a priority of the Academy and is discussed in the documents that were provided to me (Canadian Institutes of Health Research 2005, p. 6). Assessing the diffusion of research findings should probably be made one year later to allow for the normal formal and informal processes of diffusion (including publications, conferences, and grand rounds) to occur. If the Canadian Academy of Health Sciences was willing to invest in this kind of research, it could solve another problem, namely how much treatment improvements (as measured by the metrics of health care impacts) were a consequence of learning from research studies in other countries.

Space does not permit me to indicate how this dual assessment might be implemented without undue cost, but there are a variety of possible methodological solutions. At minimum, one important way of reducing the measurement costs of evaluating this aspect of ROI is to have all research studies report the percent change in the various metrics listed in Table Two for the specific morbidity involved in their project reports.

In the next section, the more familiar measures of scientific contributions (such as number of publications, patents and citations) as advances in knowledge are also suggested (see Table Four). But the above proposed measures, the advance in treatment knowledge and the increased capacity of health care personnel, are much more practical, and I would argue, fundamental ways of assessing the value of medical research. In the fourth section, I suggest how these health care impacts are translated into economic and societal benefits. It is important to keep these metrics separate because not all research results in health benefits necessarily translates into economic and societal benefits: sometimes great economic gains accrue without much health care benefit (such as in the commercialization of drugs).

Again, let me repeat that these metrics provide a great deal of flexibility in designing an evaluation of ROI because the evaluator can either sample a relatively small number of significant studies or consider a whole stream of research relative to a particular morbidity. Also, one can focus on a specific institute and the most appropriate metric for its research. Thus, the research on delivery systems can be evaluated for its contributions to the speed of diagnosis, treatment and post-treatment. Population research, and in particular, epidemiology, can be evaluated for its additions to the understanding of etiology or knowledge about a health problem. At the same time, I acknowledge that the Canadian Academy of Health Sciences is primarily concerned with macro assessment, even if it is based on meso

treatment sector metrics. Given this concern, the next topic discusses how to aggregate from the meso to the macro.

Aggregating across Treatment or Morbidity Sectors and Policy Feedbacks

Given these twenty health care impacts and the considerable flexibility they provide, how does an evaluator aggregate across disparate research findings in distinctive morbidities or treatment sectors? For example, one finding might indicate that the group method in the rehabilitation of alcoholics reduces recidivism by a certain percentage while a new back operation surgical procedure reduces side-effects and has a higher percentage of positive outcomes. This would be comparing 'apples and oranges'.

The major solution to this aggregation problem is in the computation of the percentage change, which immediately standardizes each indicator. However, this does not solve all problems of standardization. In the knowledge indicators of Table Two, the construction of indices of the knowledge base is necessary before one can compute the percentage change. This can be done crudely without undue effort. Also, there are differences between morbidity sectors in the appropriate time dimensions that are most meaningful for measuring the duration of treatment and post-treatment interventions e.g. duration of life for a week, a month, a year or reduction in the amount of hospitalization in days, weeks, etc. or speed of diagnosis, treatment and rehabilitation in minutes or hours. But in principle, these problems are solvable without giving undue weight to one particular research stream or morbidity. Here one would employ international standards as the most appropriate time dimensions.

At the same time, these percent changes need to be weighted by the caseload or relevant population in the morbidity sector, especially when computing economic and societal benefits of investments in medical research. Genetic or birth defects involve small numbers whereas cancer patients (especially in the more common cancers such as lung or breast) involve thousands. An alternative procedure is to assign weights according to the priorities of the government or the dominant values of Canadian society (Canadian Institute of Health Research 2005, p. 28). Space does not allow me to explore a variety of various kinds of weighting systems that recognize, for example, the greater difficulty in solving some problems such as the degenerative processes associated with aging. Regardless of which system is used, the key point is that weighting allows one to build in specific kinds of values about the importance of particular streams of research, another way in which policy makers can be informed.

The aggregation of potential impacts across all morbidity sectors then provides the first important assessment of the health care system, *the amount of advance in treatment knowledge*. The adjective 'treatment' is used before knowledge to help distinguish it from scientific knowledge, which is more likely to be measured by the indicators in the last section of Table Two and from contributions to knowledge that are listed in Table Four. The aggregation of the actual impacts across all morbidity sectors provides the second important assessment, *the amount of capacity added to the health care system*.

An important element in the development of any set of metrics for assessing the value of investments in research is how it informs policy makers. One of the reasons for a fine-grained conceptualization of health care impacts is that policy makers may begin to consider various trade-offs between investing in

one or another component of the treatment process. In some of the documents, interest has been expressed in research targeting. This fine-grained approach in the different morbidity sectors allows one to do this. With these indicators, one can estimate the potential pay-off from investing in a stream of research that affects a specific stage in one morbidity sector versus another stage in a different morbidity sector. Even better, the specific metrics draw attention to particular levers that health planners may want to influence (e.g. the speed of treatment or the quality of life during rehabilitation or increased effectiveness of prevention). Obviously, the potential pay-off is large from investments in research in prevention but the actual pay-offs are generally quite small because, in many cases, one cannot change behaviour that is inherently destructive (as the study of cigarette smoking, unsafe sex, and obesity demonstrate) without more research into how to make the prevention more effective. Again, the methods for changing human behaviour are beyond the scope of this white paper. Therefore, given small actual pay-offs without enormous effort and cost in attempts to change human behaviour, policy makers might prefer investments in other stages of the treatment process that appear to be more fruitful.

The treatment system and research (or knowledge production) system are quite distinct. They should be analyzed separately so that linkages between them can be better understood, which is the next topic.

Section Three: What are the Components of the Knowledge Production System? Metrics for Investments and Network Gaps

As indicated in the first section of this white paper, the discussion of investments and gaps in the way in which research is organized violates the rules of logic. Normally, one would begin with medical research investments and then proceed to the health care impacts. I have violated this rule so that the investments can be discussed in more detail, not only relative to the different morbidities, but more critically, relative to the different metrics that are listed in Table Two. By placing the metrics of health care impact first, one highlights the way in which policy makers who allocate resources have to make decisions. Admittedly, most decisions in medical research respond to the push of what researchers would like to accomplish but as the political debate increases about how funds for medical research should be allocated, the differences in expenditure on particular morbidities and especially specific metrics as objectives will be become more important. Finally, the organization of research is at the meso level where the specific impacts of Table Two can represent particular research findings more than how the entire system is organized.

One of the most important reasons to measure how medical research is organized is because the outcome will provide feedback to policy makers on how the research system can be made more effective. In particular, there has been concern expressed about how to enhance innovative capacity (Canadian Institutes of Health Research 2005, p. 8). This can only be accomplished if there is a clear image of how innovations are produced in a knowledge production system or what Hage and Hollingsworth (2000) call the Idea Innovation Network. While there has been some discussion about the utilization of research in policy making (including how to measure this kind of impact (Hanney 2007)) - and certainly influencing policy makers is one of the crucial elements in the 'pay-back' model - there is little discussion about *what kind of information should be provided to policy makers*. In other words, what do policy makers need to know? It is my contention that they need, at minimum, to know three things: (1) expenditures by morbidity, (2) expenditures relative to specific objectives defined by the metrics in Table Two, and (3) knowing how to better organize the research system relative to these

metrics. The organization of medical research is as important as how to better organize the treatment system, the objective of the Institute of Health Services and Policy. And indeed, the absence of research within this Institute relative to certain metrics may be critical for improving the speed of treatment and post-treatment interventions. Therefore, I propose a system of metrics for monitoring investments in different kinds of knowledge production defined by morbidity and metric objective within and the detection of problems in the organization of medical research that results in it being less effective than it could be for achieving the objectives outlined in Table Two.

Before discussing the different kinds of knowledge or research arenas, a definition of knowledge (one that is consistent with the definition of the treatment sector) should be provided. My definition, which is largely accepted within the sociology of science and organizational sociology is:

> *Knowledge: the sum of all the protocols involving preventive medicine, intake, intervention, and after intervention care within a certain degree of error in a particular morbidity sector as well as the base of knowledge for that morbidity.*

The advantage of this definition is that any intervention is not foolproof and comes with a certain degree of error attached to it. The real issue is to decide at what percentage one can begin to discuss knowledge as such, rather than as luck or spontaneous recovery. We have already discussed three measures of the knowledge base of the morbidity (etiology, distinction of sub-types and relevant knowledge about the physical and mental functioning of the body) in the previous section.

Research then provides advances in knowledge that in various ways improve capacity to make diagnoses, treatments, and provide better follow-up care. Within this general category of advances in medical knowledge, there are distinctive kinds of research, some of which I have already touched upon in my discussion above of the nature of the research institutes. Just as we need a fine-grained set of health care impact indicators, we also want an elaborated set of research arenas to capture the distinctive contributions of the various Canadian institutes. One such scheme is provided in the Hage and Hollingsworth's (2000) Idea Innovation Network Theory of Radical Innovation, which modified and built upon the original insights of Klein and Rosenberg (1986). The scheme identifies six ways in which knowledge advances.

Some health care policy makers might question the use of a theory developed by measuring the relationship between scientific knowledge and industrial innovation. However, the advantages of it are that it allows one to think in new ways about how medical research is organized and in particular, whether enough attention is being devoted to specific morbidities and most critically, the specific metrics within them. Transposing the names used in the industrial innovation literature to terms that are more appropriate for studying ROI from medical research, the six arenas are: basic scientific research, applied or clinical research, treatment protocol development, health care system research, quality of care research, and commercialization (including diffusion of findings research). Table Three lists three separate metrics or indicators, the first being the amount of funds invested in a specific research arena, the second the number of personnel engaged in research, and the third the number of researchers engaged in international teams. Both the second and third are measures of the training of personnel via participation in research and could easily include not only post-docs but also medical

students and even undergraduates. One of the concerns expressed by the Canadian Academy of Health Sciences (n.d.) was the building of capacity. This is the second measure of this capacity building, since the diffusion of treatment knowledge also impacts on this aspect of the health care delivery system. But rather than building capacity of health care personnel (which is what the diffusion of treatment knowledge measures) here our concern is the creation of greater research skills. The assumption is that the best training occurs while actually conducting research. If one wanted to, one could also count the number of individuals enrolled in specific training programs for particular morbidities, but this data is probably not readily available.

The four pillars of Canadian health research - Biomedical, Clinical, Health Services and Policy, and Population and Public Health - that cut across the various institutes do not exactly map on these six arenas. Nevertheless, there is still considerable overlap (Canadian Institutes of Health Research 2005, p. 21-22). Biomedical research (as defined in Canadian Institutes of Health Research 2005, p. 21) most closely fits 'basic scientific research' and includes not only studies of the eight systems of the body but also the etiology of disease and the fundamental aging processes. Likewise, Population research such as epidemiological studies can focus on this. One of the more interesting new areas is research that contextualizes treatments on the basis of genetic predispositions or other factors in the body. As I have observed, the movement of medical research, and indeed a measure of the knowledge base, is how many distinctions can be made; customization of treatments represents the extreme in this. Clinical research is also defined in CIHR (p. 21) broadly and fits the 'Applied Research' category. Health Services and Policy research is where issues about the organization of the treatment system and quality of care can be examined. Both the metrics of speed and quality can be improved via research on health services: how they should be distributed in Canada and what the best approaches with particular categories of individuals are. Finally, Population and Public Health research can also focus on problems of how best to diffuse knowledge advances given the differences in population associated with social class, gender, ethnicity, and education as well as developing more effective prevention treatments.

At the same time, these four broad categories of medical research and the mandates of the institutes do not precisely fit the different research arenas. Basic and applied research as well as protocol development can occur in almost any of the institutes. The real issue is how well connected the institutes are. If all six arenas of research are involved in any particular study, more radical protocols are developed faster.

Why develop metrics of investments in the six arenas for each morbidity or treatment sector and beyond this metric outcome? Besides the human capital implications contained in the second and third indicators within each of these arenas, there are a number of reasons as to why metrics of investments have to be part of the conceptual accounting scheme.

First, to return to the implications of these metrics for policy makers, the fine-grained approach exposes gaps in funding. This is especially true when one shifts to the ignored level of the specific metrics within the morbidity that one would like to have as objectives. For certain morbidities there may be a total absence of research on protocol development or how best to diffuse knowledge that is gained. A major desiratum of CIHR is to develop skilled researchers. This fine-grained list of six kinds of researchers also highlights lacunae that may be affecting the performance of the health care system in adverse ways.

Second, these metrics allow policy makers to consider the problem of trade-offs or the targeting of research from the other side, that is how much should be invested. However, a major issue in evaluating alternatives is the estimation of how quickly a particular objective can be achieved. These estimates are notoriously unreliable. Many scientists have searched for ways to delay the onset of Alzheimer's disease, but resolving it is an extremely difficult problem and requires an extended knowledge about the functioning of the brain during the aging process. The advantage of having twenty metrics of health care is that it provides more ways of thinking about investments in medical research beyond the broad categories of QALYs and DALYs, which are difficult to influence.

And this leads to another, and perhaps the most critical, reason to include metrics of investment: it allows evaluators to measure the social efficiency of medical research within a particular morbidity, more specifically a metric within the morbidity and by extension for the entire health care system, issues that are discussed in the next section.

Third, the third metric within each arena measures the extent of international cooperation or participation in Canadian research within an arena of a specific morbidity. International teams provide a number of advantages that are well-known: upgrading of the human capital of researchers, spreading scarce resources so that more can be learned with the same amount of investment, and possibly, making more substantial contributions to medical knowledge. The value of these collaborations is stressed by the CIHR (2005, p. 20). Their value is enhanced when one focuses on the particular metric outcome of a specific morbidity because it is so difficult for the Canadian medical research system to examine all metric outcomes in all morbidities. Indeed, as the complexity of medical research increases (in particular the movement towards customization), cooperation across national borders becomes increasingly essential. This fined-grained approach of health care metrics allows policy makers and medical researchers in Canada to select their partners with a sharper focus. For example, given the existence of a new one billion dollar centre in Germany dedicated to studying dementia, Canada might prefer having a relationship with Germany for that particular morbidity.

Table Three: Metrics of Medical Research Investments Relative to a Specific Morbidity Sector and/or Metric Outcome

Basic scientific research	Canadian dollars invested
	Number of personnel performing the research
	Number of personnel participating in international teams
Applied or clinical research	Canadian dollars invested
	Number of personnel performing the research
	Number of personnel participating in international teams
Treatment protocol development (tools, machines, techniques, procedures, etc.)	Canadian dollars invested
	Number of personnel performing the research
	Number of personnel participating in international teams
Health care system research including manufacturing of tools and equipment	Canadian dollars invested
	Number of personnel performing the research
	Number of personnel participating in international teams
Quality of health care research	Canadian dollars invested
	Number of personnel performing the research
	Number of personnel participating in international teams
Commercialization and diffusion of knowledge research	Canadian dollars invested
	Number of personnel performing the research
	Number of personnel participating in international teams

Human capital is upgraded because as the researcher learns different cognitive models associated with other national and organizational cultures, he or she develops a more complex cognitive structure of the way in which to think about a problem. Teaming with other countries with similar priorities relative to investments in medical research presumably allows for greater efficiency, that is, more advances in knowledge relative to the amount of money invested by a specific country. Finally, and perhaps most critically, teaming allows for more radical breakthroughs.

My studies of the Institut Pasteur (Hage & Mote 2007) demonstrated how complex research teams with individuals from different countries were able to make major scientific breakthroughs in biomedicine. Thus, there are many reasons to encourage international participation in medical research teams. However, these teams also have some costs attached to them, particularly if communication between different cognitive structures is reduced - an issue raised below in the discussion of metrics for gaps in the relationships between institutes*.*

Special Issues: Detecting Gaps in Medical Research by Health Care Impact and Arena of Research

As has already been suggested, two important objectives of any metric system for evaluating ROI from medical research are (1) that it provides useful information to policy makers and (2) that it represents

the 'state of the art'. Both of these objectives are achieved when the system of metrics allows one to detect gaps in medical research.

The first issue for policy researchers to ask is: are there any investments within a morbidity for each of the health care impacts listed in Table Two? Obviously, I do not have detailed knowledge of the investment portfolio of the Canadian medical research system and therefore cannot point to any particular examples of lacunae. The key point is that the fine-grained list of twenty health care metrics should allow for a careful evaluation of funding for each morbidity. In any case, when medical researchers apply for funds, it is probable that they report the expected impacts of their research on the suggested list of health care impacts.

The second question to ask is whether the six arenas are represented in a research team for a specific metric outcome. The four pillars do raise some questions as to whether enough attention is being given to developing the necessary skills for the creation of medical machines, drugs, surgical tools, and disability enhancing tools that can be commercialized, a concern of CIHR (2005, p. 17). The potential role of drug treatments to replace surgical treatments for specific morbidities, if they could be developed, would represent a large market. CIHR classifies commercialization of research as part of clinical research. In contrast, the Idea Innovation Network Theory argues this is a separate set of research skills that should be developed. Protocol development, especially for innovative products that can be commercialized, may require the participation of medical engineers, who are not normally included in the clinical category of researchers. Besides the development of new drugs, there are many other kinds of products such as hospital equipment, measurement instruments, surgical tools, etc. For example, in the rehabilitation of accidents, and especially for injuries of American soldiers in the Iraq War, there are opportunities for Canadian medical researchers to develop the special kinds of aids to walking, talking, seeing, hearing, etc. that are needed. Many of these aspects of new treatment protocols can be commercialized, although the best methods for doing so are the topic of research in the sixth and last arena. This research arena of protocol is the one most directly concerned with the problem of innovation.

Medical researchers fear that their research interests will be distorted by discussions about commercialization of medical equipment, drugs, or rehabilitation devices. Conversely, as others have observed, advances in medical knowledge are frequently driven by the interests of the researchers rather than by the priorities of the policy makers. This potential conflict is handled in two ways in the Idea Innovation Network Theory. First, specialized researchers interested in particular issues work with clinicians and basic researchers. Second, the combination of different kinds of researchers in complex teams, which are effectively integrated, increases the creativity of all, as has been demonstrated in the research of Pelz and Andrews (1976) and is the argument of Stokes (1997) about the advantages of Pasteur's Quadrant.

The fourth arena of research not only involves how best to organize the treatment process relative to a particular morbidity but also how best to manufacture new drugs and equipment. In industrial innovation, a special set of issues is the reduction of various externalities in the manufacturing of products. The same problems exist in the provision of services. The most obvious externality is the reduction of energy costs during the treatment process. CIHR (2005 p. 31) expressed some concern

about environmental impacts. It is in this arena that they can be addressed. Again, studies of this may not be occurring.

Perhaps the most distinctive arena of research is studying how best to diffuse new diagnostic tools, surgical techniques, procedures, drugs, and which ones should be commercialized. Here is an area where there is a need for a number of quite imaginative studies of how best to diffuse advances in knowledge, especially when it involves changing human behaviour, in particular in those segments of the population that are highly resistant to change.

Metrics for Knowledge Contributions

The Canadian Institutes of Health Research (2005, p. 7) has stressed the importance of outstanding research. The standard metrics for this are listed in Table Four. Rather than measuring this in the aggregate, it is important to study these contributions by morbidity sector. Again, this detailed information can be useful to policy makers who might decide that a specific morbidity sector needs more emphasis than others. These measures do not duplicate advances in treatment knowledge discussed in the previous section, especially as many basic research studies may produce few direct impacts on the treatment process and yet, of course, remain a priority of CIHR because they provide the foundation for advancing clinical research and the development of innovative protocols. Another institute that is better evaluated on the basis of scientific contributions rather than impact on the treatment processes is the Institute of Population and Public Health.

A major methods issue is the choice of the appropriate time lag between the completion of the research study and the publications attributed to it, to say little about citations referring to those publications. This is particularly complicated since the appropriate time delays for a specific study might be about three years for a publication but anywhere from five to ten years for citations after the publication appears. However, for a stream of research that slowly accumulates a body of knowledge in a specific area, one might want to use time delays that are even longer. Obviously, this makes the metrics of knowledge contributions quite different from those involving health care impacts, which can be assessed much more quickly, especially the potential impacts. This problem of time delays in the metrics of contributions to knowledge makes the advances in treatment knowledge relative to a specific morbidity a more useful feedback for policy makers because the time lags are less.

The metrics listed in Table Four measure outstanding basic research and research on service delivery, population and public health issues. In addition to the four measures that are traditionally used, I suggest that one can easily compute how much international recognition has been generated with metrics such as number of international publications and citations (whether to papers or patents). A further refinement can be counting publications in certain lead journals relevant to a particular morbidity and computing the diversity of countries represented in the international citations. Trade balances for patents are particularly critical given the Canadian government's desire for commercializing its medical research when possible. Still other possibilities for refining theses metrics are suggested by Stefan Ellenbroek (2007) in his discussion of the Leiden University's Medical Center experiences.

Table Four: Metrics for Knowledge Contributions

Metrics for a morbidity sector	Metrics for international recognition
Number of publications	**Number of international publications**
Number of citations	**Number of international citations**
Number of patents	**Trade balances for patents**
Number of patent citations	**Number of international citations**

Although prizes are an example *par excellence* of international recognition, this metric is not included. The problem with this specific indicator is the long lag time between the completion of research and its recognition by one of the major biomedical prizes (Horowitz, Lasker, Nobel Prize in Medicine and Physiology or in Chemistry, when relevant, etc.), which is usually anywhere from fifteen to thirty years.

The Metrics of Network Gaps

Above, I have suggested that one of the most important kinds of feedback for policy makers is to know how to better organize the research system. In particular, it is critical to identify gaps in funding, and even more importantly, organizational and network obstacles to doing good research, as Arnold (2004) has argued. As has been suggested, the metric system should represent the 'state of the art'. The importance of investment funds and having capabilities in all six arenas of the Idea Innovation Network Theory, especially relative to the specific metrics listed in Table Two for each morbidity, affects the amount of ROI and the speed with which new (and in particular, effective) protocols are developed. This is the basic argument of the Idea Innovation Network Theory (Hage & Hollingsworth 2000); for an example of the thinking in biotechnology and pharmaceuticals see Figure Five. If any of the links between the research organizations that handle particular arenas of research are broken or those arenas of research are not represented, radical innovation becomes quite difficult.

Figure Five: Idea Innovation Network Theory (Hage & Hollingsworth 2000)
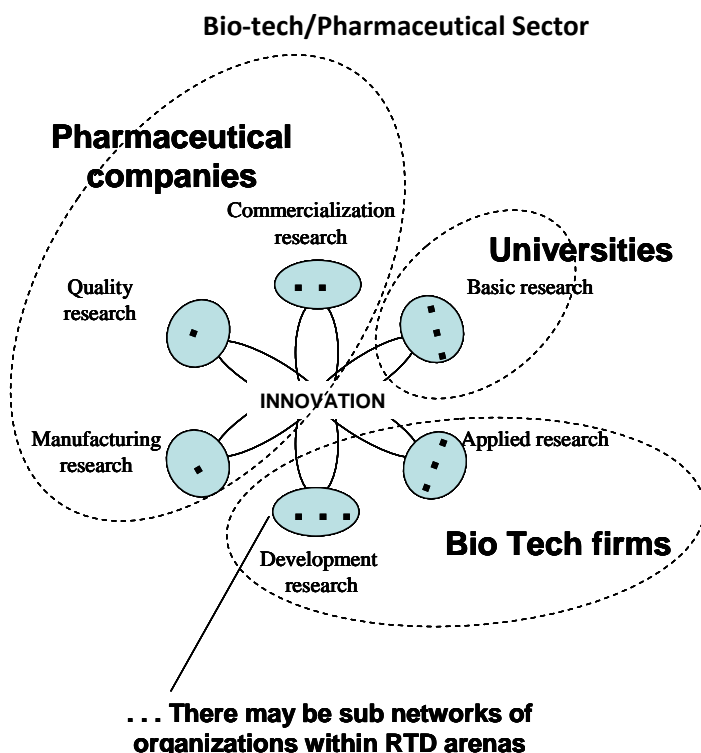
Although this example is drawn from the connection between scientific research and industrial innovation, the same logic is appropriate for medical research *per se*. For example, basic research may or may not be involved in medical schools that are attached to universities but it is certainly contained within some, if not many, of the institutes of the Canadian Institutes of Health Research. Likewise, applied research can be located both within medical schools and specific institutes. In other words, the sharp distinctions found in the relationship between scientific research, biotechnology and the pharmaceutical companies is less clear in the organization of medical research. Certainly, manufacturing research should be replaced by research on the distribution of health services, including improvements in the quality of patient care. We have already observed that the commercialization of research involves studies of how best to distribute research findings. In other words, this diagram provides some fundamental ideas that are worth consideration by policy makers and medical researchers. The search for gaps in how medical research is organized and in particular whether all six arenas of research are represented in the research team can perhaps lead to better understanding as to why Canada has been less successful in commercializing its advances in medicine than other countries.

In Figure Five, the black dots within the blue circles represent complex research teams within various pharmaceutical companies, universities, or biotech firms and by extension whether the research team is located in a medical school or one of the research institutes. The question remains though, whether basic biomedical researchers, clinical researchers, protocol development researchers, delivery of health services researchers, and diffusion of research findings experts are all represented in the same team.

The Idea Innovation Network Theory makes a number of predictions about how the Idea Innovation Network evolves with the growth in knowledge. Specifically, it argues that one needs more complex

research teams to increase the speed of radical advances in the various outcomes listed in Table Two. Creating networks that link the arenas together tightly allows for an increased rate of major breakthroughs in the development of medical protocols, including their rapid diffusion throughout the health care system. Admittedly, the objective of research is not always a major breakthrough or radical innovation because most treatments advance incrementally and frequently through experience rather than research. But it is also true that, when health care crises such as pan-epidemics occur (AIDS being one them), radical innovations and speed in the development of treatment protocols become critical.

The combination of several of these research arenas has been an important issue in the policy debates within the United States since the publication of Stokes' (1997) Pasteur's Quadrant, which argues the necessity of combining basic and applied research. The crucial issue is whether researchers with each orientation are in the same research team. Furthermore, by extension, the same argument applies to the other arenas. Indeed, it is the combination of several of these other arenas that is most likely to be fruitful. In particular, if basic and applied research is combined with protocol development and service provision, the basic and applied researchers will learn more about the problems of how the new protocol can be effectively integrated into the treatment process. The combination of protocol development with commercialization (which includes the issues of how best to diffuse the new protocol), can lead to changes in the nature of the new protocol so that it is more likely to diffuse quickly. This may mean the reorganization of the service delivery system, i.e. under these circumstances researchers who have specialized in service delivery problems also need to be included in the complex research team. This leads to the insight that one would want metrics for detecting gaps in the network that are slowing down the speed of development of medical treatments, some of which are suggested in Table Six.

The combination of all six arenas for high priority research is the major thrust of the Idea Innovation Network Theory. The most effective way of combining them is in complex research teams that include researchers from all arenas. At the same time, their integration into this quite complex research team is not self-evident. Hence the concern about not only gaps in the connections between these different arenas of research but also in the integration of the complex research teams that connect research arenas. Furthermore, the quality of the integration in the networks that connect these arenas affects the speed with which new ideas are developed. For government policy makers then, this Idea Innovation Theory allows them to look for communication problems within research teams, another gap listed in Table Six.

Table Six: Metrics for Detecting Network Gaps

Metrics for detecting network gaps between research arenas	Metrics for detecting weak linkages or communication gaps
Absence of linkages between basic and the other five	**Absence of complex research teams**
Absence of linkages between applied and remaining four	**Low communication rates within complex research teams**
Absence of linkages between protocol development and the remaining three	
Absence of linkages between service provision and the remaining two	
Absence of linkages between quality of care and 'commercialization'	

Collecting data on communication gaps inside various compositions of complex research teams may be prohibitively expensive except in those instances where the Canadian Academy of Health Sciences decides to make a particular morbidity sector a priority and is concerned about the lack of development of treatment protocols and their commercialization. Under these circumstances, Jordan's (2006) research environment survey contains a number of measures that can detect communication gaps and the absence of complex research teams. With this data, policy makers could intervene to improve the performance of their research.

Section Four: What are Returns on Investment? Economic and Societal Benefits

One of the advantages of having a fine-grained list of health care impacts is that it suggests a number of ways in which economic gains can be computed. The cost benefit studies cited by the CIHR (2005, p. 10-11) such as cardiovascular and diabetes evaluations usually cover two to five decades. However, policy decision-makers also need something much more immediate in evaluations of ROI. The advantage of having health care impacts in the treatment process by stages as well as increases in the knowledge about the morbidity is that it not only allows one to draw clearer linkages between specific research studies but it allows one to have much more immediate assessments.

In the memorandum 'A Framework to Measure the Impact of Health Research', economic impacts are not separated from societal impacts. While admitting that there is a large overlap, my recommendation (and in light of the concerns of the European Union to evaluate societal benefits as distinct from economic ones), I have tried to separate them by focusing on narrow economic benefits from various advances in metrics measuring health impacts and societal benefits by examining impacts on institutional realm performances. The logic of what I have done can, of course, be extended to include other kinds of societal performances as well.

Economic Returns

In most cases, the metrics of Table Two suggest economic gains, which are listed in Table Seven. As in Table Two, the indicators are listed under each of the four stages. But economic benefits tend to become greater the higher on the list that a health care metric has been impacted. Also, there are fewer economic metrics than health care impacts (sixteen versus twenty) because several health care impacts

can result in the same economic benefit. Also, the economic benefits of additions to knowledge about the health care problem are difficult to estimate without being able to trace directly from the knowledge to a specific health care impact. This problem is exemplified when researchers announce that someday their particular finding might result in a new vaccine or gene therapy.

Discussion is necessary along with some examples. Developing vaccines for certain morbidities has had profound impacts on the cost of treatments, the saving of lives, and the quality of life and thus there is a multiplier effect in the economic benefits. The potential economic gains via prevention are quite large even if the actual benefits in this stage of the treatment process are small when prevention requires that individuals change their behaviour. For example, a decrease in the incidence of AIDS means a considerable number of illness days saved for each individual. This value can be computed from clinical records of the average yearly cost of treatment even if this is largely a regime of drugs. The same is true for reductions in the severity of morbidity. A little exercise frequently reduces the severity of a heart attack, which in turn has a number of economic benefits. As we have seen, recent research indicates that exercise combined with a proper food regime actually changes which genes are operative and which are not.

Table Seven: Metrics of Economic Benefits from Health Care Impacts

Stage of the Treatment Process	Metric
Prevention	**Value of illness days saved from decline in morbidity incidence**
	Value of reduction in cost of treatments for less severe morbidity incidence
Intake and Assessments	**Reduction in the costs of tests for diagnosis**
	Reduction in costs of false positives or negatives
	Reduction in the costs of futile interventions
Treatment Interventions	**Reduction in the patient's costs of waiting**
	Value of life days added by successful interventions
	Value of reduction in treatment costs because of reduction in length of treatment (e.g. hospital days)
	Percent decrease in treatment costs of side-effects of intervention and/or their severity
	Percent decrease in costs of opportunistic infections during treatment intervention
	Percent decrease in treatment costs because of less invasive procedures, shift from hospital to outpatient
Post-Treatment Interventions (rehabilitation and long term care)	**Value of days saved in rehabilitation and after care**
	Percent decrease in treatment costs because of less invasive procedures, shift from rehabilitation hospital to outpatient care
	Value of increased mobility of all kinds after rehabilitation
Summary Output Measures of the Morbidity Sector	**Value of increase in the average duration of life given the morbidity**
	Value of absence of reoccurrence in health care costs and increase in the quality of life after interventions

Research on developing quick and reliable diagnostic tests can also have a considerable impact on the reduction of costs. For example, developing an effective screening technique, such as the pap smear for anal cancer in gay men not only means a quick diagnosis but it also reduces the amount of time spent in treatment and the number of life years lost when men die from anal cancer (because the detection has occurred too late for effective treatment as well as improvements in quality of life). The percent increase in the accuracy of the diagnoses reduces both false positives and false negatives. These problems are one of the major reasons why screening of the general target population is not frequently done. The costs of incorrect diagnoses are too high to justify this procedure on a cost-effectiveness basis. The accuracy of the prognosis in some of the more deadly diseases - especially at advanced stages of cancer - can lead to reductions in futile treatments. Admittedly, patients may demand them in any case. But this way of evaluating ROI of medical research is perhaps not as appreciated as much as it should be. The same can be said for certain surgical interventions (hip replacements) that have less than a 50 percent chance of success.

The next set of metrics for health care impacts on the treatment process provides a variety of ways in which economic gains can be computed. The value of the days, months, or years added to life because of successful treatment interventions (QALYs) is probably the core element of how medical research can be assessed economically. The speed of intervention can influence the likelihood of a successful outcome. Even if not, the speed reduces costs for the patient's family as they wait. The issue is: how should the value of years added to life be computed? On the basis of average salary or in terms of some stipulated value of human life? Different evaluators can make different judgments but, regardless of the decision, the logic remains the same: the economic value of the time saved. Reductions in treatment costs because of shorter hospital stays, fewer side-effects and fewer opportunistic infections are relatively straightforward and can be based on the average cost of a visit to a clinic or hospital stay per day. However, when drugs are substituted for surgery, there are issues about the costs of any side-effects from the drugs.

Given the way in which quality of life during treatment and post-treatment is computed, it means that quality also translates into certain kinds of cost-savings that are important. In particular, being able to reduce days in a hospital or in a rehabilitation hospital and most importantly in long-term care represent real and critical economic gains.

The economic gains from advances in medical knowledge about post-treatment interventions primarily apply to accidents, military injuries and degenerative processes associated with aging. The speed of intervention in these cases frequently impacts on the duration of rehabilitation and the likelihood of regaining physical and cognitive functioning. The value of days saved in rehabilitation and/or some form of institutionalized care can be based on the average cost per day for providing these services. A more difficult value to assign, but one that should be attempted, is the value of increased mobility (in the broad sense of this term) achieved by an intervention. Partial gains in walking, hand dexterity, talking, seeing, hearing, etc., mean enormous amounts to the individuals involved, even if it is difficult to assign an economic value to increased functionality. One might try to estimate the economic value by examining the likelihood of gaining employment with improvements in physical and/or mental function. This is also one of those areas where mechanical apparatuses of various kinds can be developed and commercialized.

As we indicated above, certain advances in treatment knowledge impact on more than one of these economic metrics. One should add across these different ones. For example, continuing with the example of the advantage of a pap smear to test for anal cancer: it reduces the costs associated with other tests because there is quicker intervention, which leads to reduction in costs of more invasive treatments because the diagnosis is made more quickly and can be treated more easily, and increases the value of years added to life because of quicker interventions, etc.

Finally, overall metrics, such as average increase in life expectancy relative to specific morbidity, are usually computed in cost-benefit studies of medical research such as those described by the Canadian Institutes of Health Research (2005, p. 10-13). The second metric of improved quality of life is more difficult to assess. As we have seen, some of the behavioural indicators of quality can be quantified as

economic gains but not all. The costs of living with constraints, for example, is a difficult one to assess. There are methods for doing so but these are beyond the scope of this paper. In any case, the other metrics are being emphasized as way of not having to make this assessment with surveys, one easy way in which it can be done, and instead on the basis of changes in the other metrics.

Aggregating economic benefits across treatment stages is easy within treatment sectors since the units are Canadian dollars. The problem of aggregating across morbidity sectors is the same as noted above in the discussion of the aggregation of health care impacts: how does one weight particular morbidities? Since I have already discussed this in the second section of this white paper, I merely note that this is also an issue in this aggregation procedure.

Societal Benefits

In addition to economic benefits, there are secondary societal benefits that are reflected in improved performances of various institutional sectors of Canadian society. Some of these are well-known and were discussed in the logical model analysis of the assessment of ROI. I list these as a basis for discussion and as one way of attempting to quantify the intrinsic value of medical research. Consistent with the strategy used throughout this white paper, the societal benefits would be computed for each morbidity sector involved in the assessment and then aggregated into a total benefit.

For the purposes of this exercise, I have deconstructed society into four distinctive realms: educational and scientific, political, economical, and health and welfare. There are other classification systems including my own of eight (Hage 1972) but these four reflect those that are most critical for the purposes of government policy. These institutional realms each have performance outputs. I have only suggested two but, I think, the most important two in each institutional realm.

In the educational and scientific realm, one major performance output is the upgrading of human capital of both professionals and researchers. One could construct an index of the number of new protocols learned and the number of new research projects participated in. The second output reflects the recognition of Canadian medical science. Again, one could construct an index of increased international recognition of Canadian medical science based on citations (to papers and patents) and number of foreigners that participate in research teams. The weights involved in the construction of these indices are again beyond the scope of this paper. I am merely suggesting some ways in which one could proceed.

In the political realm, the problem is to determine if medical research also involves contributing to political objectives (other than the obvious desire of the government to have improved health care). In the discussion of the importance of investing in medical research in service provision and in manufacturing of medical equipment, I suggested attempting to reduce externalities in their manufacturing, a particularly important example being the reduction of energy consumption. There are other issues involving the environment that might also be influenced by the contributions of medical research. I am not a Canadian but I am assuming that an important political objective is to maintain a good relationship with the Aboriginal community. Improved health care, and in particular, attention to how to best diffuse new knowledge to Aboriginal communities (which means investing in the

commercialization arena of research) helps build good relationships, which are important political objectives of the Canadian government.

Buxton (2007) makes a distinction between cost-savings, which are the economic benefits listed in Table Seven, and their secondary impacts on the health of the work force and on trade balances, which are included here in Table Eight. These secondary benefits of medical research are well known and need little discussion, in particular because they are mentioned by CIHR (2005, p.15). Reduction in sick days helps improve productivity as well as reduce health care insurance costs. I am not building in all of these side-benefits but in an actual assessment and with the application of logic models they could easily be included. Improved trade balances from the selling of patents and shipping of medical equipment and supplies to various parts of the world (and therefore the creation of new jobs) is a strong desire of the Canadian government. The analysis of this performance by morbidity sector may suggest some strategic areas in which to invest in the development of what I have called 'treatment protocols'.

Table Eight: Metrics of Societal Benefits from Health Care Impacts

Institutional Realm of Society	Metric
Educational and scientific performances	**Improved health care and research skill capabilities**
	Recognition of Canadian medical science internationally
Government and national performance objectives	**Reduction in energy consumption**
	Improved relations with the Aboriginal population
Economic performances	**Reduction in sick days in employment and gains in productivity**
	Increased trade balances for health equipment and supplies and creation of new organizations to manufacture health supplies
Health and welfare performances	**Increased equality in health care and duration of life by class and gender**
	Decreased pension and welfare payments as a consequence of various kinds of disabilities and aging processes

Some may be somewhat surprised that I have included the health and welfare realm since basically this whole white paper has discussed a variety of metrics involving health status. What else is there left to measure? The answer is the increased equality of health care among social classes and gender groups as well as meaningful distinctions within Canadian society. Research on the diffusion of knowledge throughout the treatment system, including how best to reach various groups with poor health, is generally a priority of governments. This is frequently the kind of contribution that the pillar of Population research can make. Although national health care systems are usually established to create equality of health care, they frequently do not, as the exhaustive study of equality in the British health

care system indicates (Hollingsworth, Hage, & Hanneman 1990). One way in which a decrease in inequality could be measured is by the reduction in the differences in duration of life.

Another important side-benefit in the improvement of health care is the reduction of pensions and welfare payments for individuals who have disabilities and/or require extended institutionalized nursing care. Sometimes governments make deliberate decisions about preferring to increase disability pensions as a way of decreasing unemployment, as the Dutch government did during the 1970s and 1980s, but the reverse process is also true. Good health care can reduce pensions because of fewer disabilities associated with the aging process. Finally, contributions to improved political performances, improved equality, and lower welfare expenditure payments might be considered as a way of quantifying the intrinsic value of medical research to society.

How does one aggregate across quite dissimilar societal performances within the same morbidity? The methodological problem of aggregation is much greater here because the units are so dissimilar. The simplest solution is to use percent change so that the units are standardized and can be combined. Even this solution poses some major problems because some of these performances are quite difficult to quantify except perhaps in surveys, e.g. the Aboriginal opinion of the Canadian government.

In practice, one can simplify these problems by focusing on only those that have economic units, especially the two measures of economic performance, and the one in health and welfare that refers to decreases in disability pensions and other programs in the welfare system that pay for individuals in long term care.

Finally, it must be admitted that this is the weakest part of the entire system of metrics I am proposing. It requires much more thought and innovative insights. But with time, I believe some interesting contributions can be made. I have only included it because I felt it appeared to be important in some of the extended discussions reported by the CIHR (2005, p. 27).

Priorities in the Selection of Metrics.

The twenty health care impacts, sixteen economic benefits, eighteen measures of research investment (not counting elaborations relative to particular health care impacts), eight contributions to knowledge, eight societal contributions, and finally seven measures of potential gaps in the Idea Innovation Network of medical research represent what might appear to be a staggering array of metrics. What should the priorities be? There are two simple criteria for selecting among these many metrics (assuming that there is not enough money to measure all of them). The first criteria is the ease of obtaining the information and the second is the amount of feedback that they provide policy makers (including researchers) interested in selecting the best opportunities for advancing Canadian health care.

On the basis of these criteria, the first fifteen health care impacts listed in Table Two should be the focus because these are easier to measure than those metrics listed under 'summary measures' and 'contributions to the knowledge background' and provide quicker feedback to policy makers. Within these first fifteen health care impacts, one could further eliminate the speed and quality measures and

perhaps also the prevention measures. This results in only seven metrics, ones that are fine-grained and can measure major incremental improvements obtained in research studies. Typically, one can ask for research projects to report exactly which of these health care outcomes has been studied. In each case, the corresponding economic gains can be relatively easily computed.

One might argue that not all research is orientated towards improving diagnoses, treatments and post-treatments. This is of course, the case. To capture the contributions to knowledge, the eight metrics in Table Four represent my next priority. Again, these are easy to measure and there are software programs, as I have already noted, that make the search for citations quite painless. Beyond this, the three contributions to the knowledge background listed in Table Two might be added but these are much more difficult to assess since they require establishing a sense of what the relevant knowledge pool is. Measuring societal contributions and the gaps in the research network would be much lower on the list of what should be measured given limited resources.

At the same time, a large range of metrics allows for policy makers to select a few metrics from the list of Table Two and then examine metrics in the other lists in light of their selection. For example, in the review of the first draft of this white paper, some preferences were expressed for measuring the quality of patient care. To respond to this request, I have indicated several different ways in which this could be assessed and I also have attempted to provide behavioural measures that are easier and less costly to measure. With this focus, one would then examine the allocation of research funds and personnel in each of the six arenas that are focusing on quality of treatment or post-treatment, whether they are finding personnel gaps in the Idea Innovation Network relative to the problem of improving the quality of life, as well as the economic and societal benefits of improvements in quality.

Aggregate Benefits as ROI from Medical Research and Computing Social Efficiency

Throughout this paper, the strategy has been to focus on a specific morbidity sector so as to better discern specific linkages between research findings, their impacts on the treatment process in at least potential benefits, and the cost savings this provides as well as societal benefits, if any. Only at the meso level can policy makers intervene to improve performance (see below). While the treatment/morbidity sector is the most appropriate one for assessments and policy interventions, it is not the best level for policy debates about the funding of medical research. For debates in Parliament, the morbidity sectors have to be aggregated to compute the ROI from medical research for Canada. At the aggregate level, one can more easily discuss how to 'improve health, longevity, and a population prepared to reach its full potential' (CIHR 2005, p. 8). But I would argue that the fine-grained approach to measuring treatment impacts would allow for the aggregation of a number of small research findings, even though there are not large impacts on QALYs. Furthermore, with the number of policy feedbacks that are provided in the system of metrics that are being suggested, policy makers can more easily introduce arguments about strategic arenas in which to invest for a larger return on medical research investment.

Given the many stakeholders that are involved, I would recommend that ROI from medical research be reported in the distinct categories outlined in the Buxton and colleagues (1994) model: (1) health care impacts; (2) knowledge impacts; (3) increased capacity of the health care system; and (4) economic benefits. The inclusion of societal benefits might be made if deemed appropriate. Health care impacts have been divided into a number of metrics. Similarly, knowledge impacts have been divided into

advances in treatment knowledge and contributions to the scientific literature. Increased capacity of the health care system involves both the increased capacity of health care personnel and of researchers. A number of economic benefits have been detailed because they are based on health care impacts.

In the discussion of investments of Canadian dollars in distinct research areas, the idea was advanced to measure the social efficiency of these investments. My definition of social efficiency is 'an improvement in an output without the assignment of dollar value to it', which distinguishes it from economic efficiency or productivity. Thus, one measures the social efficiency of the health care system at the macro level by using changes in the age-gender population pyramid divided by medical expenditures as we did in our comparison of the health care systems of Britain, France, Sweden and the United States (Hollingsworth, Hage, & Hanneman, 1990). This same logic can be applied to other kinds of health care impacts. For example, one might compute the amount of medical research dollars and researchers allocated to reduce the duration of treatment within a certain morbidity or allocated to increase the speed of post-treatment after an accident or stroke or other incident that requires rehabilitation. The advantage of having six research arenas is that it reduces the slippage between how medical research funds are spent and its consequences for the specific kinds of benefits detailed above, especially when the analysis is at the level of health care metrics.

In their cost-benefit studies, economists have stressed the estimated value of a life in the United States. In Canada, it is estimated to be 10,000 to 50,000 Canadian dollars for a QALY, 30,000 pounds in the U.K., and 100,000 dollars in the U.S. (Canadian Institutes of Health Research 2005, p. 11), and 55,000 Euros in Sweden (Robach & Carlsson 2007). Sociologists are inclined to believe the value of a Canadian life is equal to the value of an American or a Swedish life regardless of differences in salaries or cost of living. If one adopts this sociological perspective, it is better to work with a social efficiency measure that computes the percent change in QALYs in days, weeks, months, or years relative to a particular morbidity divided by the expenditures on research on that morbidity. Even broad categories of research such as cancer are best disaggregated into discrete areas because considerable strides have been made in some cancers and not in others. Again, I would suggest that this is more useful for a policy feedback to the government.

Conclusions: Four Categories of 'Pay-Back'

At the beginning of this white paper, a number of reasons were provided as to why it is important to focus on the treatment sector or meso level of analysis. One reason is the considerable variation between technologies, procedures, and expertise needed to assess different morbidities. At this level, it has become possible to specify some twenty health care metrics (and of course more could be developed) that are fined-grained enough to measure precise economic gains without necessarily having to measure QALYs, which usually are a consequence of a program of research that has unfolded in multiple countries over multiple decades. The meso level provides much more flexibility in deciding how many treatment sectors should be evaluated for determining ROI and which time periods to choose. The evaluator can assess two or three sectors or twenty or thirty depending upon the priorities and largesse of the Canadian government. Similarly, the evaluator can choose a specific research study with a lag of several years or a very long stream of research over twenty or more years, except as I have noted, one should carefully assign credit to the different countries and their researchers that were involved and evaluate the relative importance of their specific contribution, which is usually not done. Another

degree of flexibility is that the evaluation might focus on only one or two metrics and study their potential health care impacts.

Another strategic reason for the selection of the treatment sector is that it is interstitial between the micro level (where research is accomplished and patients are treated) and the macro level (where policy makers make decisions about health care policy, including which morbidity areas should receive funds). If this level is ignored, then macro policy decisions are more likely to be made without sufficient information. In particular, at this level, one can study how medical research is organized, detect gaps and blockages that prevent rapid development of radically new treatment protocols, and find leverage points where the investments of research can have the largest 'pay-back'.

I have also suggested that with the meso level, one can handle a number of important intellectual problems including different ways in which knowledge can be measured, studying organizational learning and capacity building, building social science theory, etc. But the most important reason for focusing on the treatment sector or meso level of the health care system is that it is only at this level that correct attributions can be made between research findings, their health care impacts and the subsequent economic and societal benefits that accrue.

Given the importance of the 'pay-back model' created by Buxton and colleagues (1994), the concluding remarks reflect a summary of the metrics that speak to each of the four objectives.

Knowledge Production and Capacity Building

The first category of 'pay-back' is measuring knowledge production. Three metrics were listed in Table Two to capture how knowledge about a health care problem has increased. These are separate from the approach indicated in Figure Five, where the typical metrics, papers and patents (and citations to them) are listed. Also included are some modifications in these metrics to measure the international recognition of Canadian medical science. The patents and trade balances that result from them reflect commercial knowledge production. In this regard, two of the six arenas in the Idea Innovation Network Theory of Hage and Hollingsworth (2000) also relate to this kind of knowledge, treatment protocol development and commercialization of this knowledge when it involves products such as machines, surgical tools, drugs, etc.

The metrics of Table Two measure major contributions to knowledge but these are not the only measures that are suggested in this white paper. In the second section, I made a distinction between potential benefits and actual benefits and argued that the former reflected advances in treatment knowledge, including ones that may not be associated with publications. Clinical advances frequently occur out of experience and lead to improvements that are not necessarily published. Given the variety of metrics in Table Two, some of them are sensitive to incremental improvements in the treatment process that do not affect mortality or even QALYs.

Capacity building has been treated in a number of ways. In the discussion of the actual benefits that accrue from the diffusion of treatment knowledge from new research findings, the health care

personnel's human capital is upgraded. I have also suggested that this is an interesting way to measure organizational learning. In the discussion of the investments in different arenas of research, metrics for measuring the improvement in the researchers' skills have been suggested; admittedly these are indirect metrics because they do not actually measure how much is learned. In this context, the importance of participation in international research teams has been particularly stressed because it is likely to increase the creativity of all the researchers and provide greater visibility of Canadian researchers and their achievements.

In the revised list of 'pay-back', research targeting has been eliminated, as the metrics proposed for measuring health care benefits within morbidities by stage of treatment allow for some consideration of ignored areas. When combined with the economic pay-off from specific metrics, policy makers can more easily make some judgments as to which investments might have the largest economic benefit provided that they can estimate the likelihood of achieving the objective. Finally, examining these issues across morbidities suggests which ones may not be receiving enough attention.

Informing Policy

Closely connected to the issues involving targeting are feedbacks to policy makers. Rather than developing metrics for how medical findings have influenced policy makers, I have stressed the different kinds of information that policy makers need to make intelligent decisions. The advantage of having twenty indicators of health care impact is that it considerably refines the kind of analysis that policy makers can make when searching for levers in making the biggest ROI. In some of the examples I have provided, one observes how an emphasis on speed or an emphasis on the reduction in the period of treatment or post-treatment can considerably increase economic returns because of their impact on other metrics as well.

Besides providing metrics that can be useful in targeting research, I have argued that it is important to understand the knowledge production system of medical research at each morbidity level. The six research arenas in which investments are made and skilled researchers formed may indicate some gaps in investments and in human capital. The lack of participation in international research teams may be of particular interest to policy makers because it can affect so many different aspects of the effectiveness of the health care system: the spreading of scarce resources, greater visibility of the achievements of Canadian medical research, and increased creativity in medical research.

Another set of metrics deals with network gaps between arenas and with the lack of communication within complex research teams. While this kind of data collection is expensive, the 'pay-back' of knowing how to increase the rate of scientific breakthroughs and achieve better diffusion of knowledge advances may make the cost worthwhile. In particular, this kind of evaluation can lead to better understanding of how to commercialize patentable products which flow from research and also to increase the rate of diffusion of knowledge throughout the health care system.

Just as I have argued that the detailed metrics of health care impacts allows for determining linkages with specific research studies, the list of the kinds of information feedback to policy makers will make it much easier to trace and measure when policy makers have been influenced and by how much from

medical research. Furthermore, in the justifications for budget increases, one can also trace how these arguments shift from health care policy makers to Parliament, indicating the values of the different stakeholders. These specific kinds of feedback thus become like a tracer bullet through the decision-making processes relative to medical research.

Health Care Impacts

Four issues are worth discussion in the measurement of 'pay-back' on health care impacts. First, are twenty indictors of health care impact enough? The only way to answer this is to ask if the major issues have been captured in the metrics that have been suggested. The advantage of beginning with the four stages in the process of treatment, then adding summary metrics (because these are more common in the thinking of evaluators and metrics that represent additions to knowledge about the health care problem because many kinds of medical research do not deal directly with treatments) is that this quite broad net should cover most issues. Beyond this problem, the metrics cover several major kinds of themes, such as speed, effectiveness, quality, etc. However, I suspect, as one starts coding research findings relative to a specific morbidity, new ideas will emerge as to how to assess health care impacts because they will be quite visible in the conclusions of the research study.

Second, a major distinction has been made between potential benefits, which is what can be easily coded by reading research findings, and actual benefits, which reflect the changes in the treatment process throughout the entire health care system. This distinction allows us to solve some interesting measurement problems, as we have seen. With it we can measure treatment knowledge advances as distinct from major contributions to the literature and capacity building among health care professionals.

Third, at numerous points, the methodological problem of the choice of time lags has been discussed because this plagues so much of the macro research on ROI, as Buxton (2007) and others have observed. The advantage of the metrics of health care impacts that have been proposed is that they can be based on micro time, that is, a lag of two or three years or as soon as a research project is finished. But they can also be used for macro time, that is, two or three decades, when evaluating an entire stream of research because of the incremental progress that is typical. However, for policy makers, the micro time is increasingly likely to be the operative choice.

Fourth, another methodological problem is the choice of weights. The whole strategy of the development of these metrics at the meso level is to call attention to the importance of studying specific morbidities and their treatments so as to carefully link research findings to health care impacts. But policy makers are also very much interested in the 'big picture' of what medical research has done for health. This necessitates aggregating across morbidity sectors and, once this decision is made, the problem of weighting emerges. Several suggestions as how to weight have been made but again this is a methodological issue that is beyond the scope of this white paper.

Economic Benefits

Economic benefits have been closely tied to health care impacts so that subtle economic savings can be detected. The sixteen metrics provide a good grid for capturing ROI from medical research. Furthermore, as I have noted, improvements in some of the metrics listed in Table Two have multiple economic benefits. They also more effectively highlight how research can be targeted, especially if one focuses on metrics within specific morbidities. Prevention has been discussed as having enormous potential benefits but few actual benefits (without imaginative ways of changing destructive human behaviour).

Societal benefits have been separated from economic ones because they are secondary outcomes that occur in time after initial health care impacts and their subsequent economic gains. The logic of this exercise was to distinguish institutional realms with their own performances that can be affected by advances in health status. My intent was only to open the discussion about a broader range of impacts at the macro level. Some of these might be of more interest to certain policy makers than economic gains, e.g. improvements in equality and in relationships with the Aboriginal community.

ROI would then represent the summation across all morbidity sectors weighted in various ways. But these returns are best listed separately as follows:

- Increases in knowledge about the health care problem;

- Advances in treatment knowledge;

- Enhanced capacity in treatment knowledge among health care professionals;

- Contributions to scientific knowledge;

- Enhanced capacity in skilled researchers;

- Economic gains from medical advances;

- Societal benefits from medical advances.

Since the beginning of this exercise when the development of metrics started at the meso level or the treatment/morbidity level, one can also report these more specific kinds of findings, which may be of much more interest to health care policy makers. In other words, little is lost with the extra effort that this approach entails and, as I have tried to suggest in many ways, much is gained.

References

Arnold, E., 2004. Evaluating research and innovation policy: a system world needs systems evaluations. *Research Evaluation*, 13 (1), p. 3 - 17.

Bernstein, A., et al., 2006. A framework to measure the impacts of investments in health research. In OECD (Organization for Economic Cooperation and Development), *OECD Blue Sky II Forum.* Ottawa, Canada 25 - 27 September 2006.

Borbey, P., 2007. *Measuring the impact of health research in Canada.* [Online]. Swedish Research Council. Sigtuna, Sweden: Economic Returns of Medical Research. Available at:
www.vr.se/huvudmeny/pressochnyheter/kalendarium/kalendariumarkiv/economicreturnsofmedicalresearch.5.5ec4ab911161ebf14dc80005402.html

Buxton, M., 2007. *Estimating economic benefits at the national level*. [Online]. Swedish Research Council. Sigtuna, Sweden: Economic Returns of Medical Research. Available at:
www.vr.se/huvudmeny/pressochnyheter/kalendarium/kalendariumarkiv/economicreturnsofmedicalresearch.5.5ec4ab911161ebf14dc80005402.html

Canadian Academy of Health Sciences, 2007. Prospectus for a major assessment: the return on investments in health research: defining the best metrics. [Online]. Available at:
http://www.cahs-acss.ca/e/assessments/currentprojects.php

Canadian Institutes of Health Research, 2005. Developing a CIHR Framework to Measure the Impact of Health Research. [Online]. Available at:
http://www.cihr-irsc.gc.ca/e/30418.html

Curran, B., 2007. *Impact assessment study from Irish health research.* [Online]. Swedish Research Council. Sigtuna, Sweden: Economic Returns of Medical Research. Available at:
www.vr.se/huvudmeny/pressochnyheter/kalendarium/kalendariumarkiv/economicreturnsofmedicalresearch.5.5ec4ab911161ebf14dc80005402.html

Hage, J., 1972. *Techniques and Problems of Theory Construction in Sociology.* New York: Wiley-Interscience.

Hage, J., Hollingsworth, J.R., 2000. A strategy for the analysis idea innovation networks and institutions. *Organizational Studies*, 21 (5), 971-1004.

Hage, J., Mote, J. 2007. Transformational organizations and hyper rates of innovation: the Institut Pasteur, 1888-1918. Centre for Innovation, University of Maryland (unpublished paper).

Hanney, S. et al., 2004. Proposed methods for reviewing the outcomes of health research: the impact of funding by the UK's 'Arthritis Research Campaign'. *Health Research Policy and Systems*, 2, p. 4.

Hollingsworth, J.R.,Hage, J., & Hanneman, R.A., 1990. State intervention in medical care: consequences for Britain, France, Sweden, and the United States, 1890-1970. Ithaca, New York: Cornell University Press.

Jordan, G., 2006. Factors influencing advances in basic and applied research: variation due to diversity in research profiles. In J. Hage & M. Meeus (eds.) *Innovation, science and institutional change: a research handbook*. Oxford: Oxford University Press. p. 173-195.

Kolata, G., 2008. Drug cuts risk of prostate cancer by 30%, study shows. *International Herald Tribune*, 15 June. p.4.

Lutkenhaus, J., 2008. Tinkering with acellular division. *Science*, 320 (5877), p. 755 - 756.

Marshall, E., 2008. A bruising battle over lung scans. *Science,* 320 (5876), p. 600 - 603.

Miller, G., 2008. Tackling alcoholism with drugs. *Science,* 320 (5873), p. 168 - 170.

Nelson, R. ed., 1993. *National innovation systems: a comparative analysis*. New York: Oxford University Press.

Pelz, D.C., & Andrews, F.M., 1966. Scientists in organizations: productive climates for research and development. New York: John Wiley & Sons.

Roback, K., & Carlson, P., 2007. *Economic value and innovation.* [Online]. Swedish Research Council. Sigtuna, Sweden: Economic Returns of Medical Research. Available at: www.vr.se/huvudmeny/pressochnyheter/kalendarium/kalendariumarkiv/economicreturnsofmedicalresearch.5.5ec4ab911161ebf14dc80005402.html

Stokes, D.E., 1997. *Pasteur's quadrant: basic science and technological innovation.* Washington, D.C.: The Brookings Institution.

Weisel, J.W., 2008. Enigmas of blood clot elasticity. *Science,* 320 (5875), p. 456 - 457.

Wooding, S., 2007. *The payback model – an overview.* [Online]. Swedish Research Council. Sigtuna, Sweden: Economic Returns of Medical Research. Available at: www.vr.se/huvudmeny/pressochnyheter/kalendarium/kalendariumarkiv/economicreturnsofmedicalresearch.5.5ec4ab911161ebf14dc80005402.html

Ethics and Evaluating Health Research

The Return on Investments (ROI) in Health Research: Ethical Aspects

Michael McDonald[10] and Bartha Knoppers[11]

As part of the Canadian Academy of Health Sciences (CAHS) effort to forward the project 'The Return on Investments in Health Research: Defining the Best Metrics', we have been asked to address ethical issues (Canadian Academy of Health Sciences 2007). Our aim is to provide advice to CAHS about ethically important aspects of determining the impact of investments in health research which we hereafter label as HR ROI (health research return on investment). This should be useful both to CAHS in commissioning studies in this area and to those who would carry out such studies.

In commissioning various studies (including this one) as the basis for an HR ROI 'assessment', CAHS indicates its 'general intention … to propose a clear menu of metrics by which return on investments in health research in Canada can be measured' (our emphasis). This menu of metrics would address 'the need for a robust multi-dimensional measurement framework that addresses the increasingly complex, multi-sectorial impacts of health research spanning six areas:

1. Improved health and well being

2. Benefits to the health care system

3. Improved decision making and administration

4. Creation of new knowledge

5. Increased research capacity for future innovation

6. Commercial and economic dividends (Canadian Academy of Health Sciences 2007)

We note that the six areas mentioned are central to the organizations that created CAHS. Given that the intention is to provide 'a clear menu of metrics', it is important to note both the ethical significance and the complexity of the six areas. In an ideal world health research would generate all six (Canadian Academy of Health Sciences 2007). But in reality there may be complex trade-offs amongst the goals. For example, 'commercial and economic dividends' may, in particular instances, reduce net 'benefits to the health care system' through products that produce marginal gains at significantly increased cost or reduce the capacity for the 'creation of new knowledge' by erecting major barriers to the free exchange of ideas.

In a similar paper, the Canadian Institutes of Health Research (CIHR) discusses 'developing a CIHR framework to measure the impact of health research' (Canadian Institutes of Health Research 2005). In that paper the authors note that 'where appropriate, methodologies should distinguish between social rates of return and commercial profits – innovations that have positive effects in both dimensions would

---

[10] Michael McDonald is Maurice Young Chair of Applied Ethics, W. Maurice Young Centre for Applied Ethics, University of British Columbia.

[11] Bartha Maria Knoppers is Professor of Law, Faculté de Droit, and Canada Research Chair in Law and Medicine, Centre de Recherche en Droit Public (CRDP), Université de Montréal.

be preferred to those that have negative spillovers on society'. While this seems reasonable, it is often difficult to establish when negative spillovers outweigh increased profits (other things being equal, profits are a social good) and even more complicated to weigh various social effects against each other (e.g., a health research innovation that reduces morbidity but has negative environmental effects, or the opposite – an innovation that reduces negative environmental effects at the cost of marginally increased morbidity).

In other words, there may be trade-offs that have to be made and an adequate metric would presumably improve decision-making. This raises the question of which trade-offs are ethical and which are not? Are there ethical trump cards that can be played in health research decision-making and are there ethical 'no go' areas? This leads to the topic of our paper which is on the ethical aspects of HR ROI studies.

Before beginning we offer a brief definition of ethics and ethical decision-making. From a report one of us (McDonald) co-authored, we provide the following:

> Ethics is then the science or study of morals, where morals is concerned with conduct, character, intentions, and social relations insofar as they are appraised as excellent, right, deserving, virtuous, just, or proper. In particular, ethics is concerned: with right and wrong actions, policies, and practices; with duties, obligations, and rights; with fairness in the correction of wrongs (corrective justice); with the fair distribution of benefits and burdens within society (distributive justice); with virtue, vice, and just deserts; with good and evil, benefits and disbenefits, welfare and illfare, the valuable and the disvaluable, for individuals and communities. It is concerned with the resolution of disputes, controversies, and uncertainties about the foregoing types of issues. Ethics is divided into normative ethics (opinions in morals) and meta-ethics (which is about morals), including both descriptive and theoretical ethics (McDonald 2000).

Ethical judgements are not stand-alone judgements, rather they are integrative, holistic, or 'all things considered' judgements. The Canadian moral theorist Thomas Hurka made this point well in a book on the ethics of global warming:

> An ethical judgement about climate policy is not just one judgement among many, to be weighed against economic, political, and other judgements in deciding how, all things considered, to act. It is itself an all-things-considered judgement, which takes account of economic and other factors. If a climate policy is right, it is simply right; if it is ethically wrong, it is wrong period (Hurka 1993, p. 23).

The challenge for authors and sponsors of HR ROI studies is to ensure that the studies contribute to sound 'all things considered' ethical decision-making.

Our paper takes the following form:

1. Context: Ethics in Health Research Evaluation

2. Conceptual Issues: Thinking Ethically about ROI

3. Practice Issues: Conducting Impact Assessment Ethically

4. Looking for the ROI of Ethics Research

5. Observations and Recommendations

1. Context: Ethics in Health Research Evaluation

By way of introduction, it should be mentioned that the legislation creating the Canadian Institutes of Health Research in 2000 specifically mentioned ethics in the preamble: 'Whereas Parliament believes that health research should … take into consideration ethical issues …' (Canadian Institutes of Health Research Act 2000). This historical legislation first led to the inclusion of ethics advisory boards across the Institutes, and the creation of an Ethics Office within CIHR. Yet, it is noteworthy that the 2005

'Framework for Measuring the Impact of Health Research' (Canadian Institutes of Health Research 2005) is devoid of any mention of the role of ethics in health research evaluation.[12]

Only a brief mention of ELSI (i.e., ethical, legal and social implications) is found in a table of 'indicators of health research impact and potential sources of information' (Canadian Institutes of Health Research 2005, p. 32). Indicators include the 'number of public policies influenced by ELSI 'principles'' and the 'number of clinical practice guidelines by disease area influenced by CIHR funded research' (Ibid, p. 32).

Prior to the creation of the CIHR, the Medical Research Council of Canada (MRC) had long mandated ethics review and provided guidelines to researchers since 1987 (McDonald 2000). Indeed, it was the MRC that provided the initial leadership in the creation of a Tri-Council (MRC, Social Science and Humanities Research Council (SSHRC) and the Natural Science and Engineering Research Council (NSERC)) committee to prepare a unified prescriptive statement for the ethical conduct of research involving humans to replace existing MRC and SSHRC Guidelines. This resulted in the *Tri-Council Policy Statement on the Ethical Conduct for Research Involving Humans* (*TCPS*) in 1998 (Medical Research Council of Canada 1987; Tri-Council 1998). The *Statement* was unique in that all ethics review of research involving human beings - whether social sciences, the humanities or the pure sciences - were regrouped together. While well-intended, logical and unifying, many have argued that this approach has had unintended consequences especially for the social sciences and the humanities (Maschke 2008). A further and more fundamental concern with *TCPS* (as well as other national and international guidelines and regulations in this area) is whether and to what extent they actually achieve their stated objectives of both promoting research and protecting human subjects. That is, there is a serious lack of fundamental data and performance measures for human protection (Beagan & McDonald 2005; Emmanuel et al 2004; McDonald et al 2000).

It is important to note that *TCPS* only covers research conducted at institutions that are sponsored by the Tri-Council (now known as the Tri-Agency) or in institutions that have voluntarily adopted *TCPS* as their standard for research involving humans (a number of federal agencies, the Alberta College of Physicians and Surgeons REB, and most recently Newfoundland and Labrador through legislation which has yet to be proclaimed (Health Research Ethics Authority Act 2006)[13]. However, an ever-increasing amount of health research involving humans (largely clinical trials) is sponsored by pharmaceutical companies and conducted in the private offices of physicians or through contract research organizations. Insofar as this research is part of the drug approval process it is conducted under the aegis of the *Guidelines for Good Clinical Practice of the International Conference on Harmonization* (ICH) (International Conference on Harmonisation of Technical Requirements for the Registration of Pharmaceuticals for Human Use 1997).

In health research, there are also problems as exemplified by the report of CIHR's task force on privacy (Canadian Institutes of Health Research 2005a). The other federal initiative that came to fruition after a decade of discussion and a Royal Commission is the law on assisted human reproduction and related research (Assisted Human Reproduction Act 2004). Again, the principles underscoring this legislation mention Parliament's 'ethical concerns' as justifying certain prohibitions (s.2). This legislation has far-reaching potential, well-beyond the prohibited criminal activities. Indeed, federal regulatory powers extend to the Agency created by the *Act*, which has amongst its objectives to identify ethical issues

---

[12] The same is true of the Wellcome Trust Report; nonetheless, several issues discussed in that report are ethically significant including attention to equity, shortcomings of the willingness to pay model and the like.

[13] http://www.canlii.org/nl/laws/sta/h-1.2/20080616/whole.html

(s.18(1)) and to foster the application of ethical principles (s.22). It is against this background and that of further provincial legislative and ethical overlay for biomedical research that Canada's health research community attempts to fulfill its desire to advance research and yet protect participants. Increasingly, international norms also come into play as the Canadian research community becomes part of consortia that cross borders and share data and research tools (Knoppers 2000; Dickens 2000). This latter and very recent phenomenon is not without influence on the nature and impact of ethics in health research.

This section will focus on four aspects of the role of ethics in health research evaluation: first, a discussion of the literature on the nature and role of ethics review; second, the situation in Canada; third, the confounding factor of the international nature of modern health research and finally, some comments on the lack of metrics in the role of ethics in health research evaluation.

Section 1.1: Elements for Nature & Role of Ethics Review

Since the revision of the *Declaration of Helsinki* in 2004, increasing attention has surrounded the discussion of the role of commercialization, the use of placebos, the return of research (as opposed to clinical research) results and the biobanking of tissues and DNA, especially at the level of populations (World Medical Association 2004). The latter topic, in particular where confined to sampling in defined and identifiable populations such as Aboriginal peoples, has resulted in the CIHR adopting guidelines specific to this population (Canadian Institutes of Health Research 2007).

Ethics review in Canada has also been influenced by the adoption in 2000 of the *Personal Information Protection and Electronic Documents Act* (*PIPEDA*) (S.C.), as well as the adoption of provincial legislation specific to personal information in the health care sector (Alberta Health Information Act 2000; Saskatchewan Health Information Protection Act 1999; Manitoba Personal Health Information Act 1997; Ontario Personal Health Information Protection Act). As already mentioned, the CIHR itself undertook a study of *Best Practices for Protecting Privacy in Health Research* (Canadian Institutes of Health Research 2005a). An ongoing evaluation of the contents of the *Tri-Council Policy Statement* by the Interagency Advisory Panel on Research Ethics (Interagency Advisory Panel on Research Ethics) as well as a move towards the consideration of multi-centered ethics review (Enzle & Schmaltz 2005) are influencing the role and nature of ethics review. One area of further concern is the protection of personal privacy and the transfer of data between and within institutions and provinces as well as with international collaborators (Ness 2007; Willison et al 2008).

In the decade following the adoption in 1998 of the *Tri-Council Policy Statement*, there has been an integration of multidisciplinary ethics review in both publicly and privately funded research.[14] Not only have professional societies adopted increasingly specific codes of conduct (Assisted Human reproduction Act 2004), but funding bodies (e.g., Genome Canada) themselves have insisted on the integration of ethics in applications for funding (Knoppers 2000). The same trend is appearing in the requirements for attestations of ethics review prior to publication in certain journals.[15] Publication requirements include providing proof in the case of clinical trials that: informed consent was obtained; the trial was registered; the privacy of participants was respected; Research ethics board (REB) approval

---

[14] For an example in the context of publicly funded research, see Genome Canada's International Review Panels as described in 'Guidelines and Evaluation Criteria for Position Papers', November 2007

[15] *Journal of Pharmacy & Pharmaceutical Sciences*, *Journal of Psychiatry and & Neuroscience*, *Biochemistry and Cell Biology*, *Canadian Journal of Microbiology*, *Canadian Journal of Physiology and Pharmacology*, *Applied Physiology, Nutrition and Metabolism*, *Genome*, *Journal of Palliative Care*, *Canadian Journal of Anaesthesia*, *Pain Research & Management*, *Paediatrics & Child Health*, *Canadian Respiratory Journal*, *The Canadian Journal of Gastroenterology*, *Canadian Journal of Ophthalmology*, *Canadian Journal of Nursing Research*.

was obtained; and that the *Declaration of Helsinki* was followed.[16] Universities have added internal Codes on conflicts of interest[17] to this multi-layered, complex panache of obligations and 'guidance'. However, it is not clear to what extent such codes on conflict of interest are actually observed or effective in terms of achieving their stated objective of protecting the interests of the public or the most affected groups (such as research subjects, patients, and consumers or producers of scientific information).

There are signs also of a reaction by academically based researchers against what has been called ethics review 'mission creep' (Gunsalus et al 2006). The claim is that mission creep in ethics review is endangering the system 'by excessive paperwork and expanding obligations to oversee work that poses little risks to subjects…[t]he result is that we have simultaneous overregulation and under protection'. Whether or not this latter claim is true is hard to determine given the paucity of evidence about what actually happens to research subjects and the lack of systematic study of current oversight processes. One idea worthy of consideration is commissioning HR ROI studies that assess the costs and benefits of Canadian practices and policy in key areas including conflict of interest, privacy and human research protection.

Section 1.2: The Situation in Canada

While not lamenting 'mission creep', some members of the Canadian research community have been active in decrying the quality of REB review and indeed, the governance of research generally (McDonald 2001; Jamrozik 2000; Nicholl 2000; Emmanuel, Wendler & Grady 2000). (However, it is worth noting the absence of any real participation on the part of research subjects or their advocates in debates about research protection.) Whether calling for a central independent review agency (Downie 2006) or a legal statute (Lemmens 2005) (Letendre & Lancôt 2007), the aim is to ensure greater accountability, uniformity of approach and effective oversight. There could be a centralized clearing house for decisions and supporting arguments from the letters sent by ethics review committees. This would guide future applicants, demonstrate transparency and foster natural justice through a public 'jurisprudence' of decisions rendered (McDonald et al 2000). In effect this would amount to the adoption of an administrative law approach which would recognize the fact that REB's actually act as quasi-judicial tribunals (Hadskis & Carver 2005). Indeed, this may be a welcome approach in the absence of any of the above sought after reforms since at a minimum there would be transparency in decision-making, procedural protections for members and an available 'case law' (jurisprudence) for guidance. As early as 2000, there was a call by the Law Commission of Canada for a national database of REB decisions (Beagan & McDonald 2005; Willison, Emerson & Szala-Meneok 2008).

Close to a decade later, there is finally a 'Sponsors Table' [18] advised by an expert committee[19] to look at the protection of human research participants as well as the accreditation or an alternative system of ethics review committees, and governance structures. Whether this initiative results in changes in the Canadian system of research protection remains to be seen. As noted above it would be helpful to have

---

[16] See e.g., *Journal of Pharmacy & Pharmaceutical Sciences*, 'Instructions to Authors', http://www.ualberta.ca/~csps/Journals/InstructionsToAuthors.htm

[17] See e.g., McGill University, 'A Guide to Sponsored Research', http://www.mcgill.ca/researchoffice/policies/sponsored/

[18] Human Research Participant Protection in Canada, Sponsors' Table for Human Research Participant Protection in Canada, available online: http://www.hrppc-pphrc.ca/english/sponsors.html.

[19] Human Research Participant Protection in Canada, Experts Committee for human Research Participant Protection in Canada, available online: http://www.hrppc-pphrc.ca/english/experts.html.

a serious study of the HR ROI of Canadian human research protection. In particular does it meaningfully protect subjects while promoting good health research?

Section 1.3: International Factors

Limiting ourselves to the example of human genomic research, there is no doubt that the Human Genome Project exemplified the beginning of truly international research initiatives. Founded in 1990 and culminating in 2001 with the first version of the map of the genetic sequences making up the human genome (Editor 2001), it also represented a private-public endeavour to establish a map of what has been termed pre-competitive information. Through the SNP Consortium companies, foundations, academic institutions and national funding bodies participated in this effort. Three percent of funds were officially dedicated to the study of the ethical, social and legal issues (ELSI).

Since the SNP Consortium, other international initiatives such as the HapMap Project[20] and more recently the International Cancer Genome Consortium[21] have followed this model of creating formal, international research infrastructures as resources for scientists around the world to use in their search for candidate genes and biomarkers. What distinguishes this research from traditional academic collaboration is the involvement of both the private sector and national funding bodies, the largely open-source nature of the databases, and a commitment to a common set of ethics policies and guidelines.

In spite of these common principles and practices, their implementation at the national level has not been straightforward. Historically biomedical research ethics has largely focussed on individualistic ethics, that is, respect for individual autonomy and privacy. Hence, some REB's seem uncomfortable with the 'common good(s)' and public health nature of the national and international collaborative nature of such population research infrastructures. REB evaluation has been coloured by what has been termed the 'Nuremberg' approach to the ethics review of research protocols (Rhodes 2005). The result is that creations of collaborative resources are subject to reviews unsuitable to their nature, often based on hypothetical, futuristic concerns about possible privacy invasions (Knoppers, Abdul-Rahman & Bédard 2007).

Canadian researchers involved in international clinical trials are well-served by the *Guideline for Good Clinical Practice of the International Conference on Harmonization* (International Conference on Harmonisation of Technical Requirements for the Registration of Pharmaceuticals for Human Use). But it is of note that the *Tri-Council Policy Statem*ent (Tri-Council 1998) is lacking in guidance for longitudinal, epidemiological biobanking endeavours requiring, as they do, broad consent to future research access by third party researchers for, as yet, unspecified projects. On 11 June 2008, the Canadian Partnership Against Cancer announced a project involving the creation and collaboration of cancer cohorts across Canada.[22] Moreover, CARTaGENE in Quebec is gathering genomic and health data on 50,000 randomly selected individuals[23]. Finally, the Public Population Project in Genomics (P^{3}G) is creating the tools for the international harmonization of 25 major population genomic biobanks around the world[24].

---

[20] International HapMap Project, www.hapmap.org

[21] International Cancer Genome Consortium, www.icgc.org

[22] http://www.partnershipagainstcancer.ca/inside.php?lang=EN&ID=127

[23] CARTaGENE, "Specific Objectives", http://209.44.124.200/~cartagene/index.php?option=com_content&task=view&id=13&Itemid=23

[24] Public Population Project in Genomics (P3G), http://www.p3gconsortium.org/

While the OECD is attempting to provide preliminary guidance for countries involved in such efforts (Organisation for Economic Co-operation and Development 2008), Canadian REB's (with the exception of Québec[25] and potentially Newfoundland with its new legislation(Health Research Ethics Authority Act 2006; Pullman 2005) are less prepared for the implications of the ethics of solidarity and reciprocity underlying the building of these research tools where international access and use are the norm. Likewise, retrospective access to leftover specimens from medical care or to pathological samples can be stymied by paternalistic, hypothetical concerns over individual consent (Sallée & Knoppers 2006). At the same time it is also fair to say that some international research conducted in Canada has raised legitimate concerns about the protection of the rights of donors of genetic material (Pullman 2005). In short, it may well be time to examine the guiding principles and procedures influencing Canadian practices and oversight in this area. Careful attention needs to be paid to Canadian public values, but for this to be done there needs to meaningful public consultation (Burgess 2004).

But what would be the metrics used to evaluate the return on investments in the building of open international infrastructures that serve as resources for others whether private or public? How would we ensure that these metrics reflected sound ethical values? It is to these and similar questions that we turn in the following sections.

2. Conceptual Issues: Thinking Ethically about ROI

In this section, we draw upon research in theoretical as well as applied ethics - in particular business ethics and bioethics. We begin with a conceptual exploration of HR ROI. The main idea is to determine or measure (a) the return on social investment in health research in general as compared to alternative investments a society might make (e.g., investments in health care, education, environment, industrial development, or lower taxes) or (b) the return on social investment in one form of health research as compared to another form of health research (e.g., basic biomedical research versus research on community health services). Both (a) and (b) are designed to constructively inform the formation and assessment of social policy. In principle, concepts such as ROI have general application to various types of social investment and are not limited to the health research or the health care context.

We start our exploration of HR ROI with an examination of the concept of return on investment as it is used in financial decision-making. We then take a business ethics perspective on financial investment to provide a model of ethical decision-making that will help illuminate some of the primary ethical attractions and ethical pitfalls of looking for HR ROI. We understand that some readers may initially be alienated by the application of concepts drawn from business contexts to health research and health care. However, we note that concepts such as ROI are used in various forms in private, public and not-for-profit sector planning and decision-making. We also remind readers that in fact there is a significant private sector component in health research. Moreover, the idea behind HR ROI is an attractive one, namely that of being able to make evidence-based decisions on alternative forms of social (be it public, private, or not-for-profit) investment. Whether and how ROI can be moralized or made ethical is the subject of the next sub-section.

Section 2.1 Conceptualizing ROI: Thoughts from Business Ethics

We discuss ROI in the business environment before looking at the analogical extension of the concept to health care and health research. In a widely used management accounting text, ROI is defined in the following terms:

[25] See e.g. Tubafrost: The European Human Tumor Frozen Tissue Bank, http://www.tubafrost.org/index.php.

The accrual accounting rate of return (AARR) is an accounting measure of income divided by an accounting measure of investment. It is also called the accounting rate of return or return on investment. Note that NPV (net present value), IRR (internal rate of return) and payback are based on cash flows whereas AARR is based on accrual accounting (Horngreen et al 2007, p. 825).[26]

The authors go on to note that while 'there is no uniform method of calculating AARR' the concept has some utility in helping make a choice among alternative investments. However, from an accounting perspective, other methods of evaluating alternatives (IRR and NPV) are seen by the authors as superior even though a variety of methods are used around the world (cf. Horngreen et al 2007 p. 826).

For investors and managers, ROI provides a highly useful measure of some, but not all, financial aspects of investment. In management accounting, not everything is reduced to financial measures (whether NPV, IRR or ROI). Managers and investors need to also take non-financial considerations particular to the company or organization into account. For example, whether a company's management is trustworthy and has good governance arrangements would be important for determining the long term fiscal stability of potential investments. It is also important in investment decisions to take into account external non-financial factors – such as the political stability of the relevant jurisdiction or potential changes in environmental regulation.

Thus far, we have said that prudent investors would consider both financial measures such as ROI and non-financial factors in their decision-making. We now want to add a considerable layer of complexity by considering what one would need to do to make ethical, as well as prudent, investment decisions. From a business ethics perspective, ROI and similar measures are mute on a range of ethically sensitive issues such as coercion, exploitation and oppression. These are ethically relevant and need to be taken into account by the explicit introduction of ethical criteria. Ethically or socially responsible investing is designed to do this (Boatright 1999, p. 108 ff.). However, there is some diversity on which criteria to use. Thus, if one reviews ethical investment practices, a wide range of ethical elements are invoked. Some centre on outcomes; others focus on processes. These include what the company produces (e.g., cigarettes or insulin); its modes of production (e.g., labour practices and environmental effects); its sales and purchasing practices (e.g., bribery, fairness, transparency); internal governance and accountability to shareholders; community relations including community outreach; conformity to local and international law (especially human rights charters); its political practices and a range of other factors. Some of these are seen as embodying ethical requirements (e.g., respect for employees' human rights), while others are regarded as ethically desirable (e.g., corporate philanthropy). Depending on the ethical filters selected, a given company could be regarded as either a good or bad ethical investment.

It is also worth noting that there is a dispute about the use of ethical investment criteria by those who manage pension or investment funds (Boatright 1999). Some argue that the use of such criteria is an ethically appropriate expression of universal values or alternatively that it is a reflection of the values of the particular set of investors (e.g., that university pension fund owners would not want their funds invested in companies with oppressive labour practices). Others argue that the ethical and legal responsibility of the fund manager is to maximize return on investment and that it is up to individual investors or pension-holders to decide whether they want to use other criteria for making investments. There might be a parallel debate about the use of HR ROI by private or public sector decision makers.

Whatever the merits of ROI may be for accounting and investing purposes, ROI is a financial measure designed to assess literal costs and benefits. We use the word 'literal' here because it is common in ethics and other areas of human knowledge to talk about costs and benefits in a much more extended sense that covers a wide range of non-financial aspects of processes and outcomes, e.g., mortality,

---

[26] A parenthetical explanation of accounting terms has been inserted into text.

morbidity, creativity, or political stability. This is not to deny that such extended costs and benefits have financial aspects (e.g., higher morbidity would likely result in increased health care spending and measurably decreased economic productivity). However, financial measures only capture a limited range of ethically relevant aspects unless simplifying assumptions are made that stipulate or imply the inclusion of non-financial measures or qualities that are ethically germane. For example, it could be stipulated that a disease's full cost to an individual is what that person would be willing to pay to avoid the disease. But in our view such a stipulation would only be ethically plausible provided a large range of background considerations are added. In particular, one would have to stipulate an initial fair distribution of income and other crucial opportunity factors (including bargaining power). Otherwise there would be a host of ethically problematic results, e.g., that the lives of the poor and the vulnerable would ethically count for far less than the lives of the wealthy and powerful.

We note that advocates for HR ROI studies face both practical and ethical challenges in conceptualizing HR ROI. In practical terms, ROI is easier to use and hence more attractive if it is limited to a single simple measure (e.g., morbidity). One can then compare proverbial oranges with oranges. Adding complexity to what counts as 'return' in 'return on investment' makes comparisons much harder – there is no simple index for comparison – an oranges with apples comparison. If comparisons are multi-factorial then they are potentially open to a variety of rankings depending on which factors are deemed most important. Of course, such complexity has a plus side. As we have seen in the business context a multi-factorial perspective is often more realistic. From an ethical perspective we would argue that adding this sort of complexity is essential since ethical choices generally involve context sensitive weighing of multiple factors.

This suggests two alternative approaches to using ROI in an ethically responsible way in social policy decision-making. One way would be to keep a simple measure such as NPV or ROI but hedge its use with appropriate ethical restrictions (such as justice, respect for human dignity and human rights, care for the vulnerable, and social responsibility). We label this the 'ethically constrained use of ROI'. The alternative would be to develop a multi-factorial notion of ROI appropriate to the domain of social decision-making (in particular health care) that already includes crucial ethical dimensions. This we would label as the 'ethically inclusive use of ROI'. In effect, the former produces 'other things being equal' (*ceteris paribus*) ethical judgements; whereas, the latter provides holistic 'all things considered judgements'.[27] In principle both should result in the same ethical results (assuming the same ethical values or filters at the start). However, there may be ancillary persuasive reasons for preferring one to the other in terms of how decision-makers and the public will interpret and use assessments of ROI. For example, the former (the *ceteris paribus* judgement) may be simplistically interpreted as asserting that all other things are in fact equal when they are not. This would provide a reason for using the latter ('all things considered' holistic) approach.

However, within suitable process parameters or 'side-constraints', ROI and similar results-oriented measures provide valuable tools for economic, social and political decision-making. An example of this would be a multi-faceted model (such as the 'Payback Model' proposed by Buxton and co-authors in 2004) for assessing the impact of investments in health research (Buxton, Hanney & Jones 2004). It is noteworthy that they identify equity issues as an area of sensitivity for studies that take as a metric 'benefits to the economy from a healthy workforce':

---

[27] 'All things considered' decision-making should be understood as sorting out ethically relevant from irrelevant factors. In prominent cases, some factors are deemed ethically inadmissible, e.g., race in job qualifications. Time for decision making is also taken into account. It would be ethically irresponsible if, for example, a specialist in the emergency room delayed a decision about a life and death procedure so long that the patient died anyways.

As acknowledged by many who use it, there are well-recognized problems with the human capital approach. While it tends to exaggerate benefits at a time when labour lost due to morbidity and premature mortality could easily be replaced by unemployed people or through labour migration, it limits benefits from improved health to those of working age. Thus, as a measure of the value of any health-related activity, it has uncomfortable equity implications. (Buxton, Hanney & Jones 2004, p. 736)

Equity issues could also be a factor with the three other approaches examined by authors including their favoured approach which uses economic evidence to show 'the intrinsic value to society of the health gain' from health research. In addition to equity issues, we would argue that economically based studies of HR ROI could run afoul of entitlement considerations by producing results that run contrary to pre-existing entitlements to health or other benefits from health research. If, for example, one were to argue that health research should produce the largest economic returns (however measured) per dollar that might lead to the limitation of health research in areas pertaining to small population groups (such as the military or police) that happened to have an ethical entitlement to such research (e.g., research on dealing with severe injuries from gunshot wounds).

Our suggestion would be not to refrain from commissioning such studies but rather to thoughtfully conduct the studies and soberly reflect on any ethically counter-intuitive results (as well as on results that are ethically favourable). By explicitly bringing ethics into the picture, authors and users of HR ROI studies can avoid being blindsided by ethically based objections to results and also offer a much more sensitive (and appealing) interpretation of study results.

Section 2.2: Which ethical values get built into HR ROI?

Whether we build ethics in before (ethically inclusive 'all things considered') or after (ethically restricted 'other things being equal') ROI judgements, we have to face the question of which ethical values to include. Since this is not an exercise in the construction of ethical theory from the ground up, we assume that the ethical values selected will reflect and connect with values articulated in a variety of international and national statements around human rights (including social and economic as well as civil and political rights), the rights and responsibilities of citizens in democratic societies, and the correlative obligations and rights of governments. Some of these will more particularly reflect key values in health care and health research (see Section 1 for examples of the latter). Our point here is that far from dealing with a blank slate on which we have full freedom to determine what is of value or not for assessing HR ROI, we are dealing instead with a very full slate.

However, a full slate poses challenges in its own right. Rights may conflict. So may values. As noted in Section 1, there are tensions between protection of individuals (privacy) and collective goods such as biobanks. Similarly, our society values both health impacts and economic spin-offs. How one decides to measure HR ROI in such cases reveals one's position on important value choices. The choice of such values is neither ethically neutral nor politically non-contentious.

Nonetheless, we believe that there is a reasonable degree of social consensus about central values in health care amongst Canadians. Since these have been the subject of numerous studies and reports (Romanow 2002), we shall not rehearse them here except to note that they should be taken into account in commissioning and creating HR ROI studies.

As we noted above, there are obvious pressures to find simple measures to assess the ROI in health care and health research. While these have their utility, they are not without controversy (consider the debates around QALYs – quality adjusted life years). It is important, then, to pay attention to potential problems in the use of outcome measures and the adoption of appropriate mitigating strategies. For example, there may be health research outcomes such as care or increased attention to patient autonomy that are not quantifiable in the way that morbidity and mortality are. Would this mean that health research (e.g., on palliative care) that significantly improved care or led to greater respect for

patient autonomy would therefore be assessed as less valuable than research which produced minor decreases in morbidity and mortality? Would a different choice be favoured if the gains in care were at the expense of significant gains in reduced morbidity? In regard to over-reliance on quantitative factors, the Wellcome Trust Report aptly discusses 'the seduction of numbers' (Wellcome Trust 2006, p. 33). Carefully designed HR ROI studies should be able to take into account both quantifiable and non-quantifiable factors relevant to such comparisons.

A crucial question is whether there are sound ethical reasons for setting ethical priorities on some areas of health research over others. Certainly, there have been important cases where such arguments have been made. These have been very much driven by concerns about social justice (Dresser 2001). Consider the grass roots driven initiatives to advance research into HIV-AIDS and into breast cancer. The NIH in the US has made significant efforts to put much greater emphasis on health research directed at women and at children. Various countries including the US have also targeted pharmaceutical research for rare disorders. In Canada, there has been at CIHR a major effort to direct research to Aboriginal health issues including the creation of the Institute for Aboriginal Health Research.

We believe that HR ROI studies will likely play a variety of roles in ethical priority setting for health research. One obvious role will be to see if already accepted goals are being reached. For example, is research on Aboriginal health actually addressing significant health disparities between Aboriginal and non-Aboriginal populations? Another role will be the use of HR ROI studies to shift existing priorities. Our main point here is that ethical issues should not be neglected. Instead they should be explicitly addressed in such studies themselves or in secondary studies examining the ethical premises and value assumptions made in HR ROI assessments.

Realistically HR ROI studies will be commissioned and used by those advocating investments in health research. This includes public and private sector research sponsors, researchers, research consumers and their representatives (including health charities and health coalitions). In her thoughtful work on patient advocacy and research ethics, Dresser both welcomes the growth of research advocacy (e.g., for its empowerment of patients) and offers cautionary remarks (e.g., about various forms of bias and inequality that arise when not all advocates are equally well organized and connected). Dresser suggests that attention be paid to 'ethical principles for research advocacy' (Dresser 2001, p. 159). These principles include:

1. 'First, advocates should be accurate and realistic when communicating about their work.' (p. 159)

2. 'Research advocacy should be guided by a second ethical principle: appreciation for the diversity of constituents. Like advocates, constituents are a heterogeneous group. Some constituents have the necessary self-assurance, education and economic wherewithal to be savvy research consumers. Others, however, do not.' (p. 161)

3. 'The third ethical principle is to reject parochialism in research advocacy. Advocates guided by this principle will explore the full array of policies and services that could benefit constituents. These advocates will support policies and resource allocations that advance constituents' interest in obtaining established health care as well as promising experimental interventions.' (p. 162)

In essence Dresser advocates 'responsible advocacy' which takes into account that, in an imperfect world, not all constituencies are able to voice their interests and express their legitimate concerns. Responsible advocacy is important in the commissioning, design and use of HR ROI studies.

Section 2.3 HR ROI Studies as a Work in Progress

From various sources (Buxton, Hanney & Jones 2004; Wellcome Trust 2006), we note that developing adequate metrics for assessing investments in health research is very much a work in progress. In the Wellcome Trust report on this subject, the authors conclude that:

> We conclude that there is no one 'best' method of evaluating research. Rather, various evaluation methods are complementary and different organisations and their stakeholders may employ different evaluation methods at different times. Similarly, research funders need to adopt evaluation methods that are appropriate for their research; different methods and their associated metrics need to take account of the often long, risky and incremental nature of medical research. These methods also need to recognise the value of negative findings in adding to knowledge, but also the risk that such results may be selectively under-reported. Overall, we believe there to be clear opportunities for the UK research community to develop improved evaluation methods, gain consistency in evaluation practices and demonstrate research achievements more actively Wellcome Trust 2006, p. 5).

There is not then a readymade metric that is obviously the standard of practice in this area. Rather there are a variety of metrics which capture various aspects of socially valued outputs. In developing what might be thought of as meta-metrics for work in this area, we would urge thinking holistically towards an 'all things considered' set of meta-metrics that yields results which are practically and ethically informative.

3. Practice Issues: Conducting Impact Assessment Ethically

This section involves a brief review of relevant research ethics norms likely to be relevant to HR ROI studies and an identification of some areas that may be especially sensitive. Most of these have already been introduced in Section 1. In given HR ROI studies, there may, for example, be issues of privacy and secondary use of data gathered confidentially. Similarly some HR ROI studies would require informed consent from participants (such as health researchers, health research subjects and health research consumers). Such studies would fall under the same norms and strictures as other types of research involving humans. This would involve processes and roles set out in relevant provincial, national, and international documents including the *Tri-Council Policy Statement*, provincial and federal privacy rules, and relevant foreign and international statements (for studies looking at HR ROI internationally). As noted in Section 1 there are gaps and shortcoming in the policies and practices regulating these areas, and there is significant room for improvement. Nonetheless, these are the norms and practices that are now in place. Moreover, the values that motivated these practices (respect for individual autonomy, protection from excessive research harms, the provision of benefits to research subjects) are important and worth preserving.

As noted in Section 1, some HR ROI studies target communities whose health needs have been identified as an ethical priority. In this regard, it is important to note the new *CIHR Guidelines on Research Involving Aboriginal People* (Canadian Institutes of Health Research 2007). These *Guidelines* address community as well as individual consent and are modelled on guidelines that are in use in other jurisdictions. How they will work out in the Canadian context remains to be seen. However, it is worth noting that in most regards these represent good research practices amongst experienced researchers working with Aboriginal communities.

In a different domain of research ethics, HR ROI studies are also likely to venture into areas where intellectual property and trade secrets are important factors. These need to be taken into account as valuable outputs (important, for example, to the translation of health research into efficacious pharmaceuticals). However, intellectual property and trade secrets also may constrict access to data essential for HR ROI studies.

A further area of complexity involves the likely need to make use of quality assurance and quality improvement (QA/QI) studies in determining HR ROI. These are likely to be important in working out the

long term impacts of research on practice. For example, one may want to show how a particular line of bench research improved clinical practice or how health policy research impacted health system performance. Conducting QI and QA studies in health systems or contexts will inevitably involve patients, their families and health care providers in a variety of ethically significant ways. For example, data will be gathered using patient charts or interviews with patients. Some of this data is likely to be sensitive and confidential. Health care provider practices may be observed to see whether a particular type of research was translated into clinical practice. Community members might be interviewed to see if a community participatory health research initiative made a positive difference.

The use of such QI studies raises an issue in what might be labelled ethics research bureaucratic procedures. In particular, who (if anyone) should provide ethics review for a QI? Research ethics boards are not mandated to review quality assurance, quality improvement, or performance evaluation studies (cf. *TCPS* a. 1.1d). Yet these studies may well raise the same types of ethical issues as typical research studies involving humans, including privacy and confidentiality, increased levels of risk for individuals and communities, conflicts of interest, use of human tissues and genetic materials, and the like. REBs are not generally inclined or equipped to deal with QA and QI studies. In fact the main concern for REBs has been to determine if such studies are simply sideways manoeuvres designed to escape REB review.

Fortunately there is heightened awareness around this area now and a well developed initiative in Alberta – the Alberta Research Ethics Community Consensus Initiative (ARECCI) – that has over a period of years developed and is now testing a set of useful tools and practices for conducting review in this area (Alberta Research Ethics Community Consensus Initiative 2008). We note that this initiative has been conducted in a sensitive and iterative way with a number of relevant stakeholders including researchers, research institutions, REBS, clinicians, health care institutions and administrators within Alberta and shared at two national conferences with others from across Canada elsewhere.

4. Looking for the ROI of Ethics Research

As noted in Section 1, there has been a commitment on the part of national and international research sponsors (particularly in genomics) to supporting ELSI and GE^{3}LS research. It is fair then to ask to ask about the ROI of health ethics research. One can question whether a program of health ethics research led to changes in practice and policy and then whether the changes were on the whole desirable. As in other areas of health research, comparisons and evaluations will not be easy. Sorting out multiple factors to determine the role of a specific research project in producing positive or negative effects is likely to be complicated and open to debate. Yet as noted in Section 1, there are instances where ethics research (including research in health law and the social sciences) has contributed to the creation of legislation, policy and practice provincially, nationally and internationally. One could point, for example, to the role that ethics researchers played in the Royal Commission on New Reproductive Technologies, the creation of the *Tri-Council Policy Statement*, and a wide range of other policy instruments including international and provincial norms and laws.

We have expressed concerns in Section 1 about aspects of a number of current ethics policies and practices. Whether readers share our concerns, there should be general agreement that these policies and practices do affect health research. As such they deserve critical evaluation of the sort envisaged for other types of HR ROI studies. In other words, we suggest conducting HR ROI studies of such policies and practices with an eye to possible improvements.

In most cases, examination of the ROI of ELSI research should take place in the context of more area specific studies of health research. For example, an examination of genome or stem cell research would include the contributions that ELSI research made to the area as a whole. This would still leave room for ELSI specific studies of impact in areas that are cross-cutting such as the norms for research involving

humans or animals or for the collection of health data for research purposes (rules around privacy and confidentiality of personal data).

5. Observations and Recommendations

We close with some observations and recommendations for CAHS to use in commissioning HR ROI studies. These recommendations are motivated by the idea that 'defining the <u>best</u> metrics' for the evaluation of health research involves finding a metric that is 'best' in multiple senses: accurate, illuminating, persuasive and ethical (Canadian Academy of Health Sciences 2007).

1. Thinking about the ethics of research evaluation is a new area for Canadian and international research agencies. With a few exceptions, the primary focus of these agencies has been on regulatory ethics (research involving humans and animals, conflict of interest, and intellectual property).

2. Ethically determining the return on investment for health research requires consideration of multiple normative factors that are in some cases contested. We strongly urge that these be <u>explicitly</u> articulated, discussed and defended in studies of HR ROI.

3. While quantifiable measures of health research 'investment return' are important, attention needs also to be paid to less quantifiable, qualitative indicators.

4. Among the ethical factors that the authors of this paper would defend, we include those that take into account human rights, legal entitlements, normative aspirations expressed in various national and international documents, and important contextual features. Outcomes matter ethically but so do processes. There will be tensions amongst these factors in the selection of criteria for assessing HR ROI. These should be identified and addressed.

5. Researchers conducting HR ROI studies should take into account current policies and practices that may apply to their work. These include policies for human research protection, privacy legislation, intellectual property rules, and conflict of interest. They should also be aware of changing ethical sensitivities in specific areas (such as the conduct of quality assurance and improvement) which are not currently subject to mandated forms of oversight but which are relevant to creating sound HR ROI studies.

6. HR ROI studies should look at the broad range of health research including ELSI research.

7. Impacts of existing ethical policies and practices should also be considered as important subjects for consideration as HR ROI studies.

References

Alberta Research Ethics Community Consensus Initiative (ARECCI) 2008. In Alberta Heritage Foundation for Medical Research.

Alberta Health Information Act 2000. In: R.S.A.

Assisted Human Reproduction Act 2004. In: S.C.

Beagan B., McDonald, M., 2005. Evidence-based practise of ethics review. *Health Law Review*, 13 (2-3), p. 62 – 68.

Boatright, J.R., 1999. *Ethics in finance.* Oxford: Blackwell Publisher.

Burgess, M., 2004. Public consultation in ethics: an experiment in representative ethics. *Journal of Bioethical Inquiry*, 1 (4).

Buxton, M.J., Hanney, S.R., Jones, T., 2004. Estimating the economic value to societies of the impact of health research: a critical review. *Bulletin of the World Health Organisation*, 82 (10), p. 733 – 739.

Canadian Academy of Health Sciences (CAHS), 2007. *Prospectus for a Major Assessment: the Return on Investments in Health Research: Defining the Best Metrics.*

Canadian Institutes of Health Research (CIHR), 2005. *Developing a CIHR framework to measure the impact of health research.* Ottawa: Canadian Institutes of Health Research.

Canadian Institutes of Health Research (CIHR), 2005a. *Best practises for protecting privacy in health research*: Public Works and Government Services Canada.

Canadian Institutes of Health Research (CIHR), 2007. *Guidelines for health research involving Aboriginal people.* Ottawa, 2007 May.

Canadian Institutes of Health Research Act 2000. In: S.C.

Dickens, B., 2000. Governance relations in biomedical research. In McDonald, M., ed. *The governance of health research involving human subjects.* Ottawa, Ontario: Law Commission of Canada, p. 93 – 107.

Downie, J., 2006. The Canadian agency for the oversight of research involving humans: a reform proposal. *Accountability in Research*, 13, p. 75 – 100.

Dresser, R., 2001. *When science offers salvation: patient advocacy and research ethics.* Oxford: Oxford University Press.

Editor, 2001. Human genomes, public and private. *Nature*, 409, p. 745.

Emmanuel E., et al., 2004. Oversight of human participants research: identifying problems to evaluate reform proposals. *Annals of Internal Medicine*, 141 (4), p. 282 – 291.

Emmanuel E., Wendler, D., Grady, C., 2000. What makes clinical research ethical? *Journal of the American Medical Association,* 283 (20), p. 2701 – 2711.

Enzle, M.E., Schmaltz, R., 2005. Ethics review of multi-centre clinical trials in Canada. *Health Law Review*, 13 (2 – 3), p. 51 – 57.

Gunsalus et al., 2006. Mission creep in the IRB world. *Science*, 312, p. 1441.

Hadskis, M., Carver, P., 2005. The long arm of administrative law: applying administrative law principles to research ethics boards. *Health Law Review*, 13 (2 – 3), p. 19 – 32.

Health Research Ethics Authority Act 2006. In: S.N.L.

Horngren, C.T., et al., 2007. *Cost accounting: a managerial emphasis.* Fourth Canadian ed. Toronto: Pearson Prentice Hall.

Hurka, T., 1993. Ethical principles. In H. Coward & T. Hurka, eds. *Ethics and climate change: the greenhouse effect.* Waterloo, Ontario: Wilfrid Laurier University Press.

Jamrozik, K., 2000. The case for a new system for oversight of research on human subjects. *Journal of Medical Ethics,* 26, p. 334 – 339.

Interagency Advisory Panel on Research Ethics, [Online]. Available at: http://www.pre.ethics.gc.ca

International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), 1997. *International Conference on Harmonisation of Technical Requirements for the Registration of Pharmaceuticals for Human Use: Good Clinical Practice: Consolidated Guidelines.* Ottawa, Ontario: Minister of Health.

Knoppers, B., 2000. Ethics and human research: complexity or confusion? In McDonald, M., ed. *The governance of health research involving human subjects.* Ottawa, Ontario: Law Commission of Canada, p. 109 – 115.

Knoppers, B.M., Abdul-Rahman, Mn.H., Bédard, K., 2007. Genomic databases and international collaboration. *King's Law Journal*, 18 (2), p. 291 – 311.

Lemmens, T., 2005. Federal regulation of REB review: a modest but easy step towards an accountable REB structure in Canada. *Health Law Review*, 13 (2 – 3), p. 39 – 50.

Letendre, M., Lanctôt, S., 2007. Le cadre juridique r égissant la relation entre le chercheur et le sujet de recherche: La sécurité conférée par le droit canadien et le droit québécois ets – elle illusoire? *Les Cahiers de Droit*, 48 (4), p. 579 – 633.

Nicholl, J., 2000. The ethics of research ethics committees. *British Medical Journal*, 320, p. 1217.

Manitoba Personal Health Information Act 1997. In C.C.S.M.

Maschke, K.J., 2008. Human research protections: time for regulatory reform? *Hastings Center Report*, p. 19 – 22.

McDonald, M., 2000. *Biotechnology, ethics and government: a synthesis.* Ottawa: Canadian Biotechnology Advisory Committee.

McDonald, M., et al., 2000. *The governance of health research involving human subjects.* Ottawa, Ontario: Law Commission of Canada.

McDonald, M., 2001. Canadian governance of health research involving human subjects: is anybody minding the store? *Health Law Journal*, 9, p. 1 – 21.

Medical Research Council of Canada (MRC), 1987. *Guidelines on research involving human subjects.* Ottawa: Medical Research Council of Canada.

Ness, R., 2007. Influence of the HIPAA privacy rule on health research. *JAMA,* 298 (18), p. 2164 – 2170.

Organisation for Economic Co-operation and Development (OECD), 2008. *Draft guidelines for human biobanks and genetic research databases*, Paris. [Online]. Available at: **http://www.oecd.org/dataoecd/47/28/40405918.pdf**.

Personal Information Protection and Electronic Documents Act 2000. In S.C.

Pullman, D., 2005. *Research governance, bio-politics and political will: recent lessons from Newfoundland and Labrador.* Health Law Review, 13 (2 – 3), p. 75 – 79.

Ontario Personal health Information Protection Act 2004. In S.O.

Rhodes, R., 2005. *Rethinking research ethics*. American Journal of Bioethics, 5 (1), p. 7 – 28.

Romanow, R., 2002. *Building on values: the future of health care in Canada*. Ottawa: Commission on the Future of Health Care in Canada.

Sallée, C., Knoppers, B., 2006. *Secondary research use of biological samples and data in Quebec.* Canadian Bar Review, 85, p. 137.

Saskatchewan Health Information Protection Act 1999. In S.S.

Tri-Council, 1998. *Tri-Council policy statement on the ethical conduct of research involving humans*. Ottawa: Tri-Council.

Wellcome Trust, 2006. *Medical research: assessing the benefits to society*. A report by the UK Evaluation Forum supported by the Academy of Medical Sciences, the Medical Research Council and Wellcome Trust. Independent Academy of Medical Sciences, Medical Research Council (UK), Wellcome Trust, May 2006.

Willison, D., et al., 2008. *Access to medical records for research purposes: varying perceptions across research ethics boards*. Journal of Medical Ethics, 34, p. 308 – 314.

World Medical Association (WMA), ed., 2004. *Declaration of Helsinki: ethical principles for medical research involving human subjects note of clarification on paragraph 30 added by the WMA* General Assembly. Tokyo, Japan 2004.

Public Perspective on Health Research Funding

Translating Science into Hope: The Public Perspective on Health Research Funding

André Picard

Canadians are strongly supportive of health research and see it as a worthwhile investment of both public and private monies. A survey commissioned by Research Canada showed that 85 percent of respondents believed governments should spend more on health and medical research. A similar percentage of Canadians, 83 percent, said private industry should invest more in health research and that governments should encourage them to do so with tax policies and regulations. Even more tellingly, 69 percent of those polled said they would be willing to spend out-of-pocket to support the research endeavour, albeit a modest $1-a-week (Research Canada 2006).

Yet, is public opinion polling alone enough to give us a true measure of the public pulse on this important issue? Surveys provide an important snapshot – they are an emotional litmus test across a broad range of society – of the overall mood of taxpayers. But polling has its limitations. Given Canadians' emotional entanglement with Medicare and its iconic status in the nation, asking citizens if their governments should invest more in health – be it care delivery or research – is virtually a rhetorical question. The Canadian Medical Association, in its annual survey of Canadians, routinely asks if federal and provincial governments should spend more on health delivery and research and, invariably, more than 85 percent of respondents say 'Yes' (Canadian Medical Association 2008).

Yet, asking a question without context and without explaining that cuts may have be made in other priority areas to offset increased spending gives the results less credence. In the Research Canada poll, the question about support for research was premised by the statement that only one percent of health spending goes to research, which sounds like a piddling amount. But ask the same people if $1.6-billion in tax dollars going to health research is too much and you may get a very different answer. In other words, in political polling (and what is more political than the allocation of health dollars?), how you ask the questions has a large influence on the results, and that is certainly the case in an emotionally-laden field like healthcare, and health research in particular.

Polling, because of its practical limitations, cannot and should not be the sole metric used to measure the impact and importance of health research from the perspective of the Canadian public. It must be supplemented with information from other sources, including measuring the uptake of new research findings, monitoring support for health charities, and more rigorous analysis of what health research findings capture the public imagination, and why.

The public, like the research community, is rich and varied; its views and its reasons for support of investment in health research (or not) are complex and multi-faceted. There are, in short, many publics. There is the general public that is generally disengaged and has little interest in health research until a loved one has an illness or incident that brings them into the health system; there is the informed activist public such as breast cancer survivors whose views are shaped principally by intensely personal experiences and the community of fellow survivors that has formed; there are health policy makers whose task is to translate research findings into everyday policies; there are politicians who shape the

laws and regulations that create the environment in which research is conducted and, who to a large extent, decide funding priorities; and there is the corporate public, both for-profit and not-for-profit, who pursue narrowly-focused goals, usually in one specific area of health research. Finally, there is the media, which link each of these publics in different ways and which can be a catalyst for spurring discussion on health research.

These publics often have overlapping and sometimes contradictory values, beliefs and priorities, making the singular question 'Does the public support health research?' virtually impossible to answer with a simple 'yes' or 'no'. Rather, the answer is often contextual, profoundly personal and emotional, yet not entirely immeasurable.

The Canadian Institutes for Health Research (CIHR) framework to measure the impact of health research (Organisation of Economic Co-operation and Development 2007) sets out four broad categories:

1) Advancing knowledge;

2) Informing decision-making;

3) Health impacts;

4) Economic impacts.

While the framework was developed principally with scientists and their funders in mind, each of these elements is relevant and important to the public or, more precisely, various facets of the public. However, each of these categories is likely defined and interpreted differently by scientific and non-scientific audiences and the relative weight apportioned to each category in trying to determine the overall impact of health research is likely to be significantly different between these two groups. This reality is often overlooked when searching for metrics. Just as research is multi-faceted and nuanced, so too must be the measures of its value and impact.

Advancing knowledge, or the production of knowledge, is the cornerstone of scientific endeavour. For far too long, traditional peer review has focused on outputs in terms of production of journal articles and citations and, to a lesser extent, the training of future researchers who will produce more of the same (Buxton & Hanney 1996). The public, particularly citizens with post-secondary education, places some value on the pursuit of knowledge for the sake of knowledge but, generally speaking, non-scientists do not give a whit about volume of peer-reviewed publications, impact factors and the like. These bibliometric measures have their utility but they are largely irrelevant outside academic circles, and increasingly passé for public and private funders.

Curiosity-driven research has become, in many circles, a term of derision. The feeling of much of the public is that curiosity cannot be the sole purview of the scientist; rather, exploratory research in any given field should be guided and sometimes vetted by potential users and beneficiaries, a philosophy that some scientists feels undermines their independence. Regardless of the origins of an idea, however, if scientists fail to frame research questions in a manner that is meaningful to the public, or to patients more specifically, that research will not be deemed to be worthwhile (Hoey 2002).

The growing push for accountability and transparency by funders of health research – part of a larger trend of demanding taxpayers demanding justification for spending of public funds and stakeholders holding corporations (for-profit and not-for-profit alike) accountable – has made informed decision-making a necessity. The days of supporting research projects with vague goals and indeterminate timelines are long gone. The allocation of health research dollars, and by extension scientific research itself, is no longer done in musty, cloistered laboratories, it is now a public endeavour that must be able to withstand harsh scrutiny and the vagaries of public opinion and shifting political sands. To do so, health research must not only be well done, it must be seen to be well done. The public has a deep and abiding confidence in science-based medicine but scientists must work relentlessly to maintain that confidence.

Both the supply and demand of health research is increasing, seemingly at an exponential rate (Gruman 2007). There is a broad recognition that there will never be enough research dollars to satisfy demand, though there is not always an open acknowledgement of this. Investments in health research must be strategic and targeted. The Canadian Institutes of Health Research, unlike its predecessor the Medical Research Council of Canada, is not a monolithic body. Rather, it has 13 specific institutes to help it focus its research priorities. Despite a constantly growing budget, from $260-million at its inception to $777-million today, competition for grants is fierce and relentless. The reality today is that many research projects deemed worthwhile in the traditional peer review process will only receive partial funding. (In one CIHR grants competition, in 2007, for example, 2,017 applications were received and only 331 received funding, a 16 percent success rate. Further, all the successful applicants saw their budgets cut by an average of 26 percent, and all requests for equipment funding were denied. One scientist likened the process to being asked to bake a cake and not being supplied with all the ingredients. 'You can't just bake a smaller cake,' he said (Picard 2007).)

Despite massive investments - $3.746-billion in science and technology research by the federal government alone in 2006 (Picard 2008) - the public hears, time and time again, that superb research projects are not being funded and that innovation is being stifled as a result. The situation leaves the public perplexed and angry because they have no practical way of knowing how much health research money is enough, and if they are getting value for the dollars that are invested. The current practice of selecting research to be funded based solely on scientific merit does not necessarily satisfy the needs of consumers and the broader community (Saunders 2007).

So how do policy-makers and politicians, the ultimate representatives of the public (for better or worse) in the process of allocation of health research dollars, decide what is appropriate? And how do individuals, who play an increasingly large role in the funding of health research through foundations and health charities, make their choices? Simply put, they look to get the best return on investment or, more colloquially, the best bang for their buck.

Despite the terminology, this is not strictly an economic analysis. Putting a dollar value on the benefits of health research is difficult at the best of times (Buxton & Hanney 1996). Funders of research – governments and their agencies, for-profit and not-for-profit corporations and individuals – judge the value of health research on some combination of health impacts and economic impacts. How they weight each depends on circumstance, values, philosophy and priorities.

Governments, corporate funders and universities have a tendency to favour straightforward economic measures like the commercialization of discoveries and direct cost savings to health plans and

individuals that might result from a new preventive measure, diagnostic tool or treatment. These are dollars and cents measures, often calculated over a relatively short time period and sometimes exaggerated to serve ulterior motives such as marketing of products, or pursuit of political goals.

There are, in fact, serious concerns about accountability in government-funded research. The Council of Science and Technology Advisors (2003) and the federal Auditor-General (2004) raised serious questions about accountability. More recently, we see that the post of National Science Advisor, created in part as a watchdog that would ensure that investments deliver results, was abolished in 2007. These machinations can undermine public confidence in health research.

The public, for its part, is results oriented. Yet, their interest in outcomes tends to extend well beyond monetary benefits, into more esoteric and difficult-to-measure areas. Citizens look to health research to improve their quality of life and that of their loved ones through early diagnosis, more efficient treatment and palliation, along with the economic benefits and direct cost savings that can flow from better access to care and more efficient treatments.

But the public does not always have the means to make these judgments rationally, nor an interest in doing so. Tools like QALYs (quality-adjusted life years) and DALYs (disability-adjusted life years) remain abstract academic calculators and the tradeoffs that must occur at a systems level are not always relevant to an individual in need. Levels of health and science literacy are alarmingly low. The lack of public understanding of science can create distrust, suspicion and ultimately lead to reduced funding. It also leaves the public with a lessened ability to understand and benefit from health research.

In fact, the principal criticism levelled by scientists against the public is that they are too narrowly focused on research that is immediately relevant to them and that they have unrealistic expectations. There is some truth in this, but it is a natural human reaction. A woman newly-diagnosed with ovarian cancer will have a sudden, new-found interest in ovarian cancer research; she will be scandalized to know the paltry level of research funding and clamouring for information on the latest treatments. This is not selfishness, it is rational self-interest.

It is striking too how readily many patients, even gravely ill ones, look beyond their immediate concerns and reach out to others in need. The number of Canadians who participate willingly in clinical trials for new treatments and health research projects of all sorts is staggering, and largely unrecognized. More visible are the fundraising endeavours that Canadians embrace, from the massive CIBC Run for the Cure, an annual fundraiser for breast cancer research, through to spontaneous collections that occur in communities when children are diagnosed with rare illnesses that can be treated only with obscure experimental treatments. In these instances, Canadians show their support for health research by voting with their feet and their wallets.

Fundraising is, in fact, one of the most telling metrics. Writing a cheque for a charitable donation or donning a ribbon for a cause requires more thought and commitment than simply answering a survey and suggesting governments and corporations do more. Monies raised for research by charities offer tantalizing clues about the priorities of citizens or, at the very least, reveal their top-of-mind concerns in the health research field. Yet, like public opinion polling, fundraising tallies have their limitations as a metric. In the battle for charitable dollars, cuteness counts – child-oriented charities do very well, for example – as does having a populist cause like cancer, which affects virtually every family. And the

squeaky-wheel-getting-the-grease phenomenon often results in a paradox: the best funded areas of health research tend to have it best when it comes to raising additional funds.

Regardless, the investment that health charities make in research is significant and impressive. The Princess Margaret Hospital Foundation spends $50-million annually on research and the Canadian Cancer Society $45-million a year (Picard 2008), to cite only two high-profile examples. Publicly-funded health research funding bodies would do well to take lessons from these groups on how to connect with citizens, how to address their concerns about accountability and how to track the impact of health research.

Given the abundance of choices, how do members of the public determine where to invest their health charity dollars? In other words, how does the public measure the impact of health research? What tools do ordinary citizens have at their disposal?

There is no doubt that emotion and personal experience play a central role. So too does name recognition: The Terry Fox Foundation has instant credibility because of its iconic namesake. But donors also do their homework. In the Internet age, potential sources of information are virtually limitless. The difficulty is that much of the data remain of limited value without context and the background necessary for interpretation and the public, again because of widespread scientific illiteracy, is not always able to separate the wheat from the chaff among the myriad claims in cyberworld.

Paradoxically, today, more than ever, non-scientific audiences require translators and interpreters to help decode research findings. This translation from scientific language to the plain spoken has become an essential – if not the essential – element in establishing the value of health research and in garnering public support that is increasingly central for the launch and continuance of health research projects.

Funding agencies like CIHR have wisely invested heavily in knowledge translation. They have made it central to their core mission. The CIHR's mandate is the 'creation of new knowledge' and 'its translation into improved health for Canadians'. The latter is not optional, it is integral. And this translation of new knowledge into better health outcomes cannot occur in a vacuum. The necessity of public engagement in the health research endeavour is implicit and for the public to be engaged it must be informed.

Health charities, foundations that fund health research and corporate bodies like pharmaceutical companies have understood this reality for far longer than public funders. They were leaders in knowledge translation long before the term became voguish. The breadth and depth of some of these sources is staggering and their ability to engage the public is impressive. So too is their influence, which they have leveraged through the careful cultivation of relationships with the media. There are few institutions in modern society that garner as much attention – not to mention as much positive attention – as health charities like the Heart and Stroke Foundation of Canada and the Arthritis Society of Canada.

Health charities have succeeded in bolstering support for health research, and for research in their areas of interest specifically, with well-thought-out and carefully executed strategies. First, a cause is humanized by telling heart-wrenching personal tales, such as surviving a bout with cancer. The stories are not mere tear-jerkers; they are bolstered with statistics, such as the staggering number of cancer diagnoses yearly, and they tend to focus on a specific public policy issue, such as the wait time for a particular surgical intervention. These stories, covered extensively by the media, generate a tremendous

outpouring of sympathy along with financial donations which, in turn, are invested in research. The researchers, who benefit from the funding, are thus engaged in the cause and become trusted and valued spokespeople. They also generate new findings and new products that engage corporations. This, in turn, lends itself to new campaigns, more media stories, successful fundraising, and the cycle continues.

Public funding bodies have long taken the tack that research should speak for itself, an approach that engages neither the public nor the media, and leaves much worthy research far from the public eye. Publication in a peer-reviewed journal and presentations at scientific conferences alone can no longer be considered adequate dissemination of research findings. It is barely the beginning of the process. Of late, however, public funding bodies have come to emulate health charities and they do a lot more outreach and formal public relations. They have also come to recognize that vignettes and anecdotes can be used to successfully engage and educate the public, along with raising the profile of research projects and researchers.

Case studies and narratives allow for the easy-to-understand demonstrations of policies, products and clinical practice changes that can engage the public. Ultimately, the public understands the value of health research not on a global basis, but based on case-by-case exposure to research findings that, they come to see, cumulatively, as worthwhile and relevant to themselves and their loved ones (Mollas-Gallar 2000).

But the starting point is knowledge translation, the presentation of research findings in digestible morsels. Albert Einstein, arguably the greatest scientist of all time, said it best: 'Most of the fundamental ideas of science are essentially simple, and may, as a rule, be expressed in a language comprehensible to everyone'. Scientific bodies, and scientists themselves, who ignore or pay short shrift to knowledge translation do so at their peril.

Yet, there can be no question that the cultures of scientific and non-scientific writing are very different, and that can blur and confuse the message. Scientific writing, as it appears in traditional, peer-reviewed journals, is cautious, stilted and impersonal. The language and terminology are often impenetrable to the general public, and deliberately so. There is an emphasis on methodology and a reluctance to bring the findings to life by underscoring their practical utility.

Non-scientific writing about science and health, on the other hand, aims to be broadly accessible. It relies heavily on anecdotes and personalities, to the point where the science and the findings of the scientific research may seem almost secondary. Journalists, contrary to researchers, often have little hesitation to take even the most obtuse, theoretical findings and speculate wildly on their boundless possibilities. Thus, in the mainstream press, the 'cure' for cancer is commonplace, leaving the ever-hopeful but constantly disappointed public mistrustful and cynical. Despite the stereotyping and the excesses at either extreme of the spectrum, however, much scientific journalism is accurate and informative.

A survey of researchers who published their research findings in major journals like *Science* and *Nature* showed that 87 percent were pleased with the media coverage of their work. A similar study, published in the *Canadian Medical Association Journal*, found that 11 percent of articles about genetic research contained information that was 'exaggerated', implying that the vast majority of coverage was highly accurate (Bubela & Caulfield 2004).

What is clear is that the public gets the vast majority of their information about healthcare and health-related research from the media, principally from television and increasingly from the Web. What is less clear is what the public does with these nuggets of health information and how they influence their perception of health research and the view they hold of the value of this investment.

The true value of health research, ultimately, should be judged by its ability to positively influence behaviour and improve health outcomes. But the path from publication of research findings to altered health behaviour is rarely linear. Research findings pass through many filters and there are countless other influences, on an individual, familial and societal scale. Practically, cultural and behavioural change can take years; even physicians can be slow to implement guideline-compliant care, or to act on compelling, widely disseminated research findings.

There is anecdotal evidence, however, that the process of influencing health behaviours through research findings can be greatly accelerated through knowledge translation, particularly when media afford extensive coverage to a research finding and influential health charities weigh in. Two recent, striking examples of this phenomenon are the Women's Health Initiative (WHI) finding that hormone replacement therapy (HRT) can increase, rather than decrease the risk of heart attack and stroke in post-menopausal women (Roussow et al 2002) and the finding that the popular painkiller Vioxx increases the risk of cardiovascular events (Graham et al 2005). Within months of the publication of the WHI findings sales of HRT, one of the top-selling categories of prescription drugs plummeted; the reaction was not due solely to the research article in the *Journal of the American Medical Association* – which virtually no one in the general public actually read – but driven by the vocal concerns expressed by users and women's health groups. In the case of Vioxx, regulators acted with unprecedented swiftness – even before the damning research findings were actually published in a peer-reviewed journal – and a $10-billion-a-year prescription drug was rendered close to worthless, in large part because of the unrelenting attention on the drug by the traditional media and new media like Web sites and bloggers.

Scientists and their funding bodies need to recognize that the public seems to react more quickly and unequivocally to these informal sources of advice than to the underlying scientific findings that are invariably more nuanced. This should not unduly alarm researchers but, rather, serve as a reminder of the value and necessity of knowledge translation in a timely and efficient manner. Health researchers and their funders must accept that, if they want to have an impact on the public's health, publication of rigorous scientific findings is, in itself, not sufficient. There is a need for researchers to descend from their ivory towers and engage in the discussions that take place in the offices of general practitioners, around kitchen tables, and in cyberspace chat rooms because, ultimately, this is where the most impactful and lasting health decisions are made. This too is where citizens' views on health research are shaped, and scientists should not be disengaged spectators in this process.

Many of us were born into an era where we took for granted antibiotics, vaccines, CAT scanners and microsurgery and knew little or nothing about tuberculosis, diphtheria, typhoid and smallpox (Clement 2008). Health research has altered the course of history and the state of the world. The contribution of research to the early detection, treatment and prevention of disease, not to mention extending life expectancy and bolstering quality of life has been remarkable (Peipert 2002). The reduction of health-care costs that has resulted from these advances demonstrates that research is a wise investment. However, the scientific challenges that remain are considerable and unquantifiable.

Equally challenging is communicating advances to the public. The reality is that most scientific advances are incremental. Today, there are few 'home runs' like the discovery of the polio vaccine that changed the lives of millions virtually overnight. Still, these tiny steps are essential in the journey to better health and we should not lose sight of this reality in our impatient society. The risk exists that, in the quest to demonstrate success, scientists and funding bodies will aim for easily measurable impacts. This will almost certainly lead to minimal and short-term gains and be self-defeating over the long-term.

Investing in health research is an economic gamble (Grens 2007) but it is a gamble that invariably pays off over time. A leap of faith is required from the public and there is no shame in saying so. But that faith needs to be bolstered through open communication and promotion of understandable examples of scientific progress. If only a tiny fraction of research results in people living longer and healthier, the resulting cost savings will mean payoffs will be many times the initial investment (Joint Economic Committee of the United States Congress 2000). That it is impossible to predict precisely which research will pay off most is an unfortunate reality. Similarly, there needs to be recognition that there is no magic formula for determining the 'right' investment in health research.

There is currently a popular formula that holds that one percent of health spending should go to research. The figure was not arrived at by some complex scientific formula; it is an arbitrary number, based as much on whimsy as anything else. That does not make it an invalid goal. Rather, it serves as a useful reminder that how big the health research pie should be and how it is sliced is largely a political process. That process needs to be more democratic and more transparent. It needs to intimately involve consumers of health services and citizens more broadly. The more engaged they are in the promise of science, the more informed they are about the potential and the limitations of health research, the better. The public should have high expectations because health research has delivered on a grand scale.

There is every indication that the public is strongly supportive of health research through their words and actions, but that support should not be taken for granted. Investments in health research must be justifiable, and justified, not only to review committees of peers, but to a larger public.

Ultimately, there is no single means of measuring the public pulse on investing in health research, there is no unique appraisal instrument, no single metric that can be agreed upon (Grens 2007). The level of investment in health research, and the distribution of those research dollars, should not be a mere mathematical calculation but a collective expression of the desire to be healthy, an expression of hope on a grand scale.

References

Bubela, T.M., Caulfield, T.A., 2004. Do the print media 'hype' genetic research? A comparison of news stories and peer-reviewed research papers. *Canadian Medical Association Journal*, 170 (9), p. 1399 – 1407.

Buxton, M.J., Hanney, S.R., 1996. How can payback from health services research be assessed? *Journal of Health Services Research Policy* 1 (1), p. 35 – 43.

Canadian Medical Association, 2008. *8th Annual national report card on health care*, [online]. Available at:
http://www.cma.ca/multimedia/CMA/Content_Images/Inside_cma/Annual_Meeting/2008/GC_Bulletin/National_Report_Card_EN.pdf

Clement, T., 2008. [Address to awards ceremony of Gairdner Awards], 5 May, 2008.

Graham, D., et al., 2005. Risk of acute myocardial infarction and sudden cardiac death in patients treated with cyclo-oxygenase 2 selective and non-selective non-steroidal anti-inflammatory drugs: nested case-control study. *The Lancet*, 365 (9458), p. 475 – 481.

Grens, K., 2007. An economic gamble: what does society get for the billions it spends on science? *The Scientist*, 21 (7), p. 28.

Gruman, J., 2007. *Communicating health research in the era of headline news.* [Address to Media Science Forum], 6 December, 2007, Ottawa.

Hoey, J. ed., 2002. The epistemology of epidemiology. *Canadian Medical Association Journal*, 166 (2), p. 157.

Joint Economic Committee, United States Congress, 2000. The benefits of medical research and the role of the NIH. [Online]. Available at: http://hsc.utoledo.edu/research/nih_research_benefits.pdf

Mollas – Gallar, J., et al., 2000. Assessing the non-academic impact of grant-funded socio-economic research: results from a pilot study. *Science and Public Policy*, 9 (3), p. 171 -182.

OECD (Organisation for Economic Co-operation and Development), 2006. *Proceedings of Blue Sky II Forum 2006; Science, technology and innovation indicators in a changing world: responding to policy needs*. Paris, France 2006.

Peipert, J.F., 2002. The economic value of medical research: is it worth the investment? *Obstetrics & Gynecology,* 99, p. 835 – 840.

Picard, A., 2007. Slim budgets blunt cutting-edge research. *The Globe and Mail*, 1 Feb., p. A13.

Picard, A., 2008. Five cents of every research dollar goes to cancer research – is it enough? *The Globe and Mail*, 26 Aug., p. L4.

Research Canada, 2006. *Canada Speaks 2006: Research Canada's first public opinion survey on health research in Canada*. [Online]. Available at: http://www.rc-rc.ca/en/content.php?doc=52%20.

Rossouw, J.E., et al., 2002. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized control trial. *Journal of the American Medical Association,* 288 (3), p. 321 – 323.

Saunders, C., et al., 2007. Beyond scientific rigour: funding cancer research of public value. *Health Policy*, 84 (2 – 3), p. 234 – 242.

Health Research Evaluation Frameworks: An International Comparison

Philipp-Bastian Brutscher

Steven Wooding

Jonathan Grant

TR-629-DH

August 2008

Preface

This report is based upon, and summarizes findings from eight research evaluation frameworks in use in the UK, Sweden, the US (2), the Netherlands, Australia, the EU, Canada and elsewhere. This report was jointly supported by the Canadian Academy of Health Sciences (CAHS) and the International Observatory on Health Research Systems. The Observatory is funded by the Health Research and Development Policy Research Unit of the UK Department of Health.

The CAHS has convened an Assessment Panel to consider what research evaluation framework would be most appropriate in a Canadian context; and to look at what modifications might be needed to such a framework to adapt it for the Canadian context. The objective of the present study is to inform the work of the Panel by providing an overview and comparison of international research evaluation frameworks.

The report is divided into two parts. In the first part, five key elements of research evaluation (emerging from the frameworks studied) are presented and discussed: evaluation objectives, outcome measures, levels of aggregation, timing and evaluation methods. In addition, correlation diagrammes are used to explore the relation between these elements. The second part presents case studies on the eight evaluation frameworks studied.

The report is based on desk-based document review and key informant interviews. The report will be of interest to government officials dealing with health and medical research policy, medical research councils, health and medical research charities, public and private institutions engaged in health research, and researchers.

RAND Europe is an independent private, not-for-profit, research institution that helps improve policy and decision-making through research and analysis.[28] For more information about RAND Europe or this document, please contact:

Dr. Jonathan Grant
RAND Europe
Westbrook Centre, Milton Road
Cambridge CB4 1YG
United Kingdom
Email: jgrant@rand.org
Tel: +44 (0) 1223 353329

Dr. Steven Wooding
RAND Europe
Westbrook Centre, Milton Road
Cambridge CB4 1YG
United Kingdom
Email: wooding@rand.org
Tel: + 44 (0) 1223 353329

---

[28] For more information on RAND Europe, please see our web site: www.randeurope.org

Acknowledgments

The authors are grateful for the interest and assistance of all those who contributed to this report. In particular, they would like to acknowledge the valuable input on the research evaluation frameworks studied, provided by the following individuals:

Dr Stefan Ellenbroek (Leiden University Medical Center), Dr Marcus Nicol (Australian National Health & Medical Council), Dr Linda Butler (National Australian University), Dr Johan Froeberg (Swedish Research Council), Dr David Cox (UK Medical Research Council), Julie Tam (UK Department of Innovation, Universities and Skills) and Cpt. Melissa Kame (US Congressionally Directed Medical Research Programs).

Within RAND Europe, thanks for ongoing thoughts and feedback is due to Helen Rebecca Schindler and Observatory team members Amanda Scoggins and Dr Tom Ling. The briefing has furthermore benefited from the comments of Quality Assurance reviewers Dr Sonja Marjanovic and Dr Tom Ling.

Abbreviations and terms

CAHS	Canadian Academy of Health Sciences
CDMRP	Congressionally Directed Medical Research Programs (US)
CoA	European Court of Auditors
CWTS	Centre for Science and Technology Assessment, Leiden University
DG	Directorate General
DIUS	Department of Innovation, Universities and Skills (UK)
GPRA	Government Performance Results Act
HERG	Health Economic Research Group (Brunel University, UK)
LUMC	Leiden University Medical Center
MRC	Medical Research Council (UK)
MORIA	Measure Of Research Impact and Achievement
NIH	National Institute of Health (US)
OMB	Office of Management and Budget (US)
PART	Program Assessment Rating Tool
RORA	Record of Research Achievement
RQF	Research Quality Framework
USMRMC	United States Army Medical Research and Materiel Command
ZonMW	The Netherlands Organization for Health Research and Development

Executive Summary

The creation of new knowledge and its translation into innovation does not occur overnight. The underlying processes are complex and characterized by challenges revolving around (among other things) the ability to appropriate the returns to investment in research and asymmetric information (e.g. between researchers and research funders).

It is often argued that, as a consequence, there is a role for public policy with regard to supporting research and its translation into innovation.[29] Moreover, there is an increasingly prevalent view that evaluation can play a crucial role in this context.[30] It can: help to overcome problems of "asymmetric information"; provide a better understanding of results flowing from policy interventions; allow learning from past experiences; and provide elements for improving strategy definition.

More specifically, in this report we identify and discuss four rationales for research evaluation. We argue that research evaluation (if well designed and implemented) provides the ability to: 1) hold researchers, funding bodies and/or policy-makers better accountable for their action; 2) "steer" research (into a desired direction); 3) "signal" ability (on the part of researchers, for example to show that they are worth funding); and 4) provide input into the research management process (helping to improve strategy definition etc).

The main part of the report is based upon, and compares, eight international research evaluation frameworks in use: the Leiden University Medical Center (LUMC) framework; MORIA; PART; the Vinnova; Payback and UK Department of Innovation Universities and Skills (DIUS) frameworks and the frameworks of the European Union and the Congressionally Directed Medical Research Programs. The frameworks were identified on the basis of desk research and chosen in discussion with the Chair of the CAHS Panel.[31]

On the basis of these frameworks, in a first step, we identify and discuss five key elements of research evaluation frameworks:

- Evaluation objectives, which flow from the four rationales of evaluation outlined above: accountability; "steering"; signalling; and advocacy;

- Outcome measures, ranging from output measures, comprising the goods and services directly produced to impact measures, capturing the long-term changes research brings about;

- Levels of aggregation, which may be low (in case of an individual researcher, for example), intermediate (in case of a faculty or research programme) or high (when a whole research discipline is evaluated);

- Timing, which can be cross-sectional (if an evaluator is interested in the outcomes of one piece of research) or longitudinal (if the evaluator is interested in the outcomes

---

[29] Fahrenkrog, G. et al (2002): *RTD Evaluation Tool Box – Assessing the Socio-Economic Impact of RTD – Policy*; IPTS Technical Report Series.

[30] Boehkolt, P. (2002): *Innovation Policy and Sustainable Development: Can Innovation Incentives make a difference?*; IWT Observatory

[31] Other frameworks can be found in Hanney et al. (2007): *An Assessment of the Impact of the NHS Health Technology Assessment Programme*; Health Technology Assessment; 11(53)

from a research group over a certain period of time, for example, rather than a particular piece of research); and

- Evaluation methods, comprising statistical data analyses, modelling methods (such as microeconometric modelling) and qualitative and semi-quantitative methods (such as interviews and case studies).

Comparing the evaluation frameworks we studied along these five key elements we find that the frameworks differ significantly: The payback framework, for example, has an accountability objective, output measures, a low level of aggregation, a short (longitudinal) time frame and is based on a handful of qualitative and semi-quantitative methods. The DIUS framework, on the other hand, has a "learning" objective, impact measures, a high level of aggregation, a cross-sectional time frame and a whole plethora of evaluation methods it draws upon.

In a next step, we look at the interdependencies of these key elements. We examine to what extent an evaluator or policy maker faces trade-offs between the choices he or she makes with regard to different key elements. That is, we look if the choice of an accountability objective for example has any bearing on the choice of an outcome measure. This question is highly relevant from an evaluator's and/or policy-maker's perspective, because (if such a trade-off exists), this suggests that there are better (and worse) combinations of key elements and that a careful (rather than ad hoc) examination of the choice of these elements is crucial.

We suggest that, from a theoretical perspective, it is likely that such trade-offs exist. In addition, we use correlation diagrammes (based on the frameworks studied) to further explore these tade-offs. The small sample size of eight frameworks does not allow us to come to a definitive answer. Yet, we find some evidence in the direction that trade-offs exist:

- Accountability and advocacy objectives, we find, tend to be associated with "upstream measures" (i.e. outputs/outcomes), whereas "steering" and "learning" objectives tend to be associated with "downstream measures" (i.e. outcomes/impacts).

- Upstream measures, in turn, we find, tend to be associated with low levels of aggregation, whereas downstream measures tend to be associated with high levels of aggregation.

- Similarly, upstream measures tend to be associated with shorter evaluation intervals (in case of longitudinal evaluations), whereas downstream measures with longer intervals.

- Low levels of aggregation, we find, tend to be associated with fewer evaluation methods, whereas high levels with more methods.

From this a second conclusion follows: trade-offs in the choice of key elements of evaluation frameworks are likely to exist. As a consequence, key elements should be chosen very carefully – taking into account that elements which appear appropriate in isolation need not be a good choice when combined with other key elements.

In particular, the choice of an evaluation objective, we find, is immensely important. It, directly or indirectly, influences the appropriateness of all other key elements.

Further empirical research is required, however, to base this conclusion on a more robust basis.

Introduction

Government officials and business representatives constantly stress the importance of research for the economy. It is seen as a main input into the innovation process, a contributor to growth, employment and international competitiveness, and a source of prestige. There is also the social aspect: innovations flowing from research help people to live longer and in better health, they help to preserve the environment and to make life easier for people, giving them more free time and more ways to spend it.[32]

Yet, advances in research do not occur overnight, even less so their translation into innovative products and services. The underlying processes are complex and characterized by a number of market failures.[33] As a consequence, it is often argued that "a clear commitment and bold forward-looking strategy [for supporting research advancement and its translation into innovations] on the part of policy makers [and research funders] is needed".[34]

There is an increasingly prevalent view that evaluation can play a crucial role in this context.[35] Polt et al (2002), for example, find that: "[i]ncrease in the complexity and uncertainty present in policy decision-making requires the emergence of strategic intelligence combining the synergies of capacities between evaluation, technology foresight and technology assessment, to produce objective, politically unbiased, independent information to support active decision-making."[36]

In fact, as shall be argued in the following, evaluation (if well designed and implemented) can help to reduce problems of market failure, provide a better understanding of results flowing from policy interventions, allow learning from past experiences and provide elements for improving strategy definition.

This report is based upon, and summarizes findings from eight research evaluation frameworks in use in the UK, Sweden, the US (2), the Netherlands, Australia, the EU, Canada and elsewhere.[37] It is divided into two main sections. The first section provides a synthesis of key findings of the eight frameworks. The second section gives a summary of each framework.

Rationale for R&D support by governments

Government support for research is typically justified on the grounds of market failure. The idea is that under some circumstances free markets result in an inefficient resource allocation.[38] There are a

---

[32] Witt, U. (1996): "Innovations, externalities and the problem of economic progress" in: *Public Choice*; Vol. 89; pp.113–130

[33] Metcalfe, J.S. (2003): "Equilibrium and Evolutionary Foundations of Competition and Technology Policy: New Perspectives on the Division of Labour and the Innovation Process"; in: *Revista Brasileira de Inovacao*; Vol.2; No.1; pp. 112–146

[34] Fahrenkrog, G. et al (2002): *RTD Evaluation Tool Box – Assessing the Socio-Economic Impact of RTD – Policy*; IPTS Technical Report Series p.13

[35] Boehkolt, P. (2002): *Innovation Policy and Sustainable Development: Can Innovation Incentives make a difference?*; IWT Observatory

[36] Fahrenkrog, G. et al (2002): *RTD Evaluation Tool Box – Assessing the Socio-Economic Impact of RTD – Policy*; IPTS Technical Report Series.

[37] In addition to Canada, the Payback framework has been applied in a number of countries – see case study for an overview.

[38] By efficiency we mean Pareto efficiency. An allocation is Pareto-efficient if no individual can be made better off without making another individual worse off.

number of reasons why in the context of research, markets are likely to "fail".[39] Two of the most prominent ones are "knowledge spillovers" and "asymmetric information".

As research is (to a large extent) concerned with the production of new knowledge, this leads to what are known as "knowledge spillovers". According to this concept, because of the "public good" properties of knowledge[40] (and acknowledging that intellectual property rights influence the extent to which knowledge is a public good and the types of knowledge that are considered such), the benefits from research do not accrue to the research performer only, but "spill over" to other individuals, firms, industries, even economies.

That is, because of the "public good" properties of knowledge, individual researchers (as well as firms, industries or economies) can benefit from activities undertaken by others for (almost)[41] no cost – i.e. without having to replicate those activities internally. As a consequence, researchers are likely to hold back their efforts (to some extent), hoping to benefit from the efforts undertaken by others.[42] From a society's perspective, this implies that investment in research is likely to be too low (relative to the Pareto optimal yardstick) and that markets "fail".[43]

Knowledge spillovers have often been taken as an argument for (strengthening) intellectual property rights.[44] In addition, because this remains insufficient, they have also been taken as an argument for public funding of research.[45] Intellectual property may not be sufficient (to deal with the problem of knowledge spillovers) because, as Griliches (1990) argues, not all knowledge can be protected by intellectual property rights.[46] Moreover, even if it can, Scotchmer (1991) claims that it is often difficult to define the right breadth and scope of intellectual property (to efficiently deal with spillovers).[47]

"Asymmetric information" describes the situation in which an imbalance of knowledge exists between parties – for example between researchers and potential suppliers of capital. That is, potential lenders

---

[39] See Arrow (1962): "Economic welfare and the allocation of resources for invention"; in R.R. Nelson (ed), *The Rate and Direction of Inventive Activity: Economic and Social Factors*; pp. 609–626, NBER

[40] By public goods properties we mean that codified knowledge is neither excludable nor rivalrous. That is, no one can be effectively excluded from using it and its use by one individual does not reduce the amount of knowledge available for use by others.

[41] Cohen, W.M. et al. (1990): *Absorptive Capacity: A New Perspective on Learning and Innovation* suggest that, in order to benefit from research efforts undertaken by others, individuals (firms, industries, economies) have to invest in research themselves (hence do incur "costs"). For a formal presentation of this point see: Leahy D.; Neary, P. (1997): "Public Policy Towards R&D in Oligopolistic Industries"; in: *The American Economic Review*; Vol.87; No.4; pp.642–662

[42] This argument follows from the assumptions made in Rational Choice Theory and is typically referred to as the "free-rider problem" – see for example Metcalfe, J.S. (2003): "Equilibrium and Evolutionary Foundations of Competition and Technology Policy: New Perspectives on the Division of Labour and the Innovation Process"; in: *Revista Brasileira de Inovacao*; Vol.2; No.1; pp. 112–146

[43] Nelson, R. et al (1982): *An Evolutionary Theory of Economic Change*; Cambridge, MA; Harvard University Press.

[44] Ibid – Intellectual property can reduce the effect of spillovers by granting the inventing researcher the sole right to use his or her invention.

[45] Nelson, R.R. (1959): *The Simple Economics of Basic Scientific Research*; University of Chicago Press

[46] Griliches, Z. (1990): *Patent Statistics as Economic Indicators: A Survey*; Journal of Economic Literature, 28(4); No.4.; pp. 1661–1707

[47] Scotchmer (1991): *Standing on the Shoulders of Giants: Cumulative Research and the Patent Law*; Journal of Economic Perspectives; Vol.5; No.1; Other reasons include that intellectual property (in some situations) hampers diffusion; that it can have anti-competitive effects and also that it can lead to "patent races". – see for example Clark, and/or D. and M. Blumenthal (2007) "Rethinking the design of the Internet: The end to end arguments vs. the brave new world" *TPRC*, Arlington Virginia

sometimes cannot accurately judge the credibility of claims made by researchers/research groups.[48] Problems of "adverse selections" and, in particular, "moral hazard" are a consequence, both of which can work to decrease the incentive to invest in research, causing (as well) an inefficient allocation of resources.[49]

"Adverse selection" refers to the situation in which, due to informational asymmetries (or other factors), a higher number of less-qualified researchers tend to apply for and receive R&D funding than otherwise.[50] "Moral hazard" describes the problem of people not bearing the full consequences of their actions (under asymmetric information) and consequently behaving differently (e.g. showing less effort) than they would if what they were doing was perfectly observable.[51] One way to deal with problems of asymmetric information (as we shall argue) is evaluation.

Rationale for R&D evaluation

Evaluation can be defined as "a systematic and objective process designed to assess [ex post] the relevance, efficiency and effectiveness of policies, programmes and projects".[52]

There are four broad rationales for R&D evaluation:[53] 1) to increase accountability (of researchers, policy-makers and funding bodies), 2) to "steer" the research process, 3) to provide a means for "advocacy" (for researchers/research groups), and 4) to provide an input into the management process (through better understanding and learning).

The first rationale follows directly from the problems of "asymmetric information": A systematic evaluation of research (capturing outputs, outcomes and impacts) provides a measure (albeit imperfect) of researcher activity. This, it can be argued, increases visibility and the possibility to hold researchers accountable for their behaviour, reducing problems of "adverse selection" and "moral hazard".

As an example, if a funder for medical research wants to make sure her money is used productively by a researcher, she can either monitor the researcher closely or evaluate her (on the basis of the outputs, outcomes and impacts she produces). Choosing the latter, the research funder can use the findings of the evaluation (such as a very low research output) to make inferences about the behaviour/activity of the researcher (taking into account other possible explanations for the findings).

However, not only does the behaviour of researchers become more transparent through evaluation, but also that of funding bodies and policy-makers. To the extent that outputs, outcomes and impacts can (also) serve as an imperfect measure of the behaviour of funding bodies and policy-makers, evaluation (also) increases visibility of their behaviour and the possibility to hold them accountable for it.

---

[48] Stoneman, P., Vickers, J. (1988): "The Assessment: The Economics of Technology Policy"; in: *Oxford Review of Economic Policy*; Vol. 4; No.4; pp. I–XVI

[49] Laffont, J.J. et al (2002): *The Theory of Incentives: The Principal–Agent Model*; Princeton University Press

[50] Akerlof, G. (1970): "The Market for Lemons: Quality Uncertainty and the Market Mechanism"; *Quarterly Journal of Economics*; 84(3)

[51] Laffont, J.J. et al (2002): *The Theory of Incentives: The Principal-Agent Model*; Princeton University Press

[52] Fahrenkrog, G. et al (2002): *RTD Evaluation Tool Box – Assessing the Socio-Economic Impact of RTD – Policy*; IPTS Technical Report Series

[53] For an alternative (more narrow) list see: Georghiou, L. et al (2005): "Evaluation of Publicly Funded Research; Report on the Berlin Workshop"; downloaded from: www.internationales-buero.de/_media/Report_on_Evaluation_Workshop.pdf

If, for example, the funder of medical research (from above) repeatedly fails to allocate its funds productively (and to fund research that results in the discovery of new molecules, for example), then (in the absence of other explanations) he may be held accountable for this failure.

The second rationale for evaluation, which is an increased ability to steer the research process towards desired outcomes, goes hand in hand with the idea of increased accountability. The reason is that evaluation does not only make research activity more transparent but allows (to some extent, at least) for researchers to be "contracted" in a way that maximizes the chances of producing what is desired (in terms of outputs, outcomes and impacts).

As an example, if the same funder of medical research is interested in a specific achievement, say the discovery of a new molecule, (rather than only the productive use of his money in general) then he can set (ex ante) a target to discover a new molecule for the researcher, and use evaluation (ex post) to check if the target has been achieved (and to hold the researcher accountable, if this is not the case) thereby "steering" the research process (towards the discovery of a new molecule).

Not only can the activity of researchers be "steered" but also that of policy-makers and funding bodies. As an example, if a policy-maker is interested in the discovery of a new molecule he can (just as the research funder in the example before) set (ex ante) a target to discover the molecule for research funders (rather than researchers), "contract" them, and use evaluation (ex post) to check if the target has been achieved.

The third rationale for research evaluation is the flip side of the first one (i.e. to use evaluation to "screen" for information on researcher, policy-maker or funding body behaviour). The idea is that often researchers (policy-makers or funding bodies) have an interest to "signal" their ability to conduct research (or to fund it). Evaluation can be used to do so (acknowledging (positive) past performance). This rationale can be referred to as "advocacy".

Finally, it has been argued that evaluation of research can help to understand policy results better and allow for learning from past experience. This provides elements for improving strategy definition, resulting in increased efficiency and efficacy of policy interventions. As Polt et al. argue: "Evaluation tools have expanded to provide […] means […] to facilitate mutual learning from past experiences, supporting mediation, decision-making and policy strategy definition."[54]

Background to the study

The objective of the present study is to inform the work of the Panel convened by the Canadian Academy of Health Sciences (CAHS) by providing an overview and comparison of international research evaluation frameworks. First, on the basis of desk research, 12 international research evaluation frameworks were identified. In discussion with the Chair of the CAHS Panel, 8 (of the 12) frameworks were selected for further analysis (the LUMC framework, MORIA, PART, the Vinnova, Payback and DIUS frameworks and the frameworks of the EU, and the CDMRP). For a summary, see table below (Table 1).

The main focus for the selection was to balance the degree of novelty of the frameworks and the context in which they are used (such as basic and applied research). See figure below (Figure 1). The

---

[54] Fahrenkrog, G. et al (2002): *RTD Evaluation Tool Box – Assessing the Socio-Economic Impact of RTD – Policy*; IPTS Technical Report Series p.13

slight bias towards more recent evaluation frameworks can be explained by the momentum research evaluation work has gained over the last decade or so.
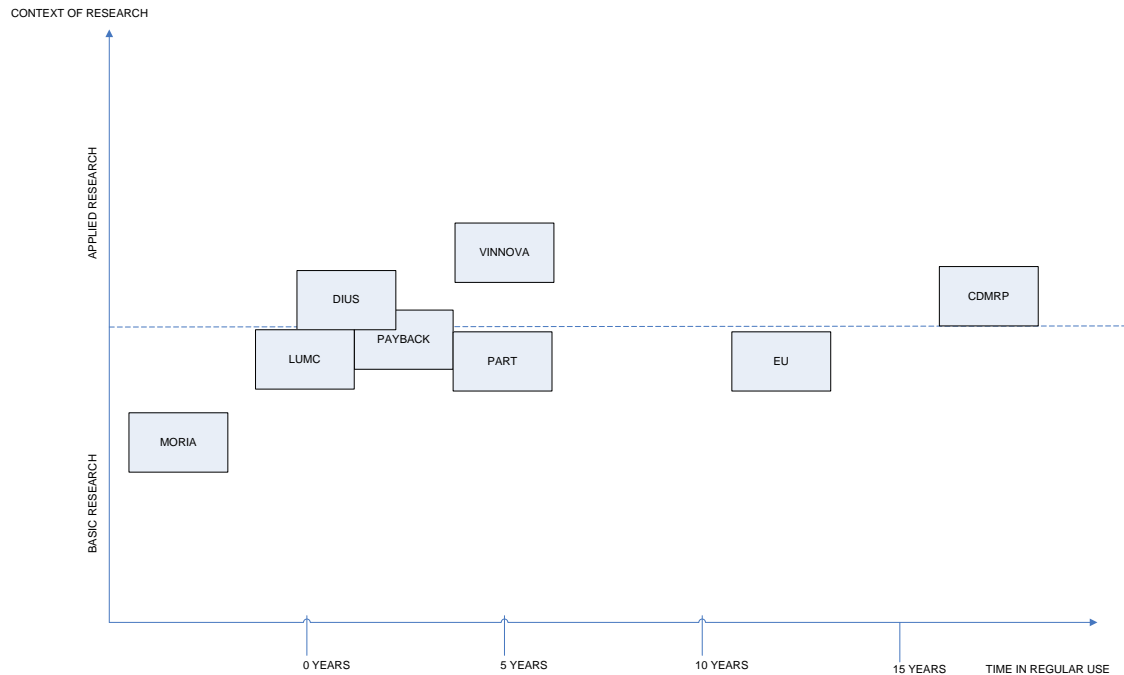
Figure 1 Research Evaluation Frameworks studied – by type and time in use

On the basis of the initial search, a case study template was developed. The idea of the template was to ensure that similar and comparable information would be collected for each framework. The template was reviewed by the Chair of the CAHS Panel to ensure that all areas of interest to the Panel were covered.

On the basis of the common understanding and agreement achieved through the template review, the RAND Europe team then completed the case studies. These were based on desk research and, where practical, email contact and telephone interviews with key informants in the organizations selected. [55] To ensure that all information was correct, after completion the case studies were sent (back) to individuals in the respective organizations. [56]

In a final step, the findings from the case studies were analysed in a RAND Europe internal workshop. The results were then written up and quality assured.

---

[55] Many thanks to Stefan Ellenbroek, Marcus Nicol, Johan Froeberg, David Cox, Julie Tam, Cpt. Kame.
[56] Except for PART and the CDMRP (For Vinnova: Johan Froeberg)

Frameworks	Country	Description
Leiden University Medical Center (LUMC)	NL	The framework in place at the Leiden University Medical Center (LUMC) is an ex post evaluation framework which focuses on the "societal impact" of research at the level of the research group. Looking at "societal impact" (rather than scientific quality), the framework can be seen as part of a broader movement in the Netherlands to correct for the "serious imbalance in the research portfolio" (arising from a sole focus traditionally of evaluation on scientific quality).[57]

The underlying assumption of the framework is that societal impact and scientific quality need not always go hand in hand. Smith explains: "Much research that scientists judge of high quality has no measurable impact on health – often because the lag between the research and any impact may be decades. Thus scientists would think of the original work on apoptosis (programmed cell death) as high quality, but 30 years after it was discovered there has been no measurable impact on health. In contrast, research that is unlikely to be judged as high quality by scientists – say, on the cost effectiveness of different incontinence pads – may have immediate and important social benefits".[58] |
| Measure of Research Impact and Achievement (MORIA) | AUS | MORIA stands for "Measure Of Research Impact and Achievement".[59] It looks at outputs, outcomes and impacts of research across three domains: "knowledge", "health gain" and "economic benefits". MORIA was developed at the Australian NHMRC as an analytic (support) instrument in the (ex ante) peer review process for grant applications. It builds on the Record of Research Achievement (RORA) framework. At the moment, it seems unlikely that MORIA will be used in this (ex ante evaluation) function. Some of the work may, however, be used in the NHMRC post grant assessment.

A particularly interesting aspect of MORIA is its scoring system. Similar to the LUMC framework, findings are translated into a (standardized) numerical score. This allows comparison and aggregation of findings across projects and (within projects) across different domains. |

---

[57] Smith, R. (2001): "Measuring the Social Impact of Research"; *BMJ*; 323; pp.528
[58] Ibid p.529
[59] NHMRC (2006): "National Health and Medical Research Council Submission to the Productivity Commission's Research Study on Public Support for Science and Innovation in Australia". Downloadable from: http://www.pc.gov.au/__data/assets/pdf_file/0013/38110/sub080.pdf (accessed on 18.8.2008)

Frameworks	Country	Description
Program Assessment Rating Tool (PART)	US	PART stands for "Program Assessment Rating Tool". It was introduced shortly after George W. Bush took office in 2001, as part of his agenda to improve government management. PART is used to assess the effectiveness of around 800 federal programmes. It takes the form of a "diagnostic questionnaire". An interesting element of PART is that (to a large extent) it evaluates programmes on the basis of performance goals. To do so, it adopts output, outcome and efficiency measures. Most weight is on outcome measures.
Vinnova (Swedish Governmental Agency for innovation systems)	S	Vinnova is the Swedish Governmental Agency for innovation systems. When Vinnova was formed in 2001, there was an interest in understanding better what its initiatives were achieving, as well as in developing methods to estimate its long-term impacts. Since 2003, Vinnova has been conducting impact analyses of its work on a yearly basis. The Vinnova framework consists of two main parts: an ongoing evaluation process and an impact analysis. There is some variation in how the framework is applied. The discussion in this report is based on the recent work on traffic safety.
Payback (in use at the Canadian Institute of Health Research)	CA	The Payback framework was developed at the Health Economic Research Group at Brunel University (HERG). It has been applied in a number of different contexts. (It has been used by, for example, the UK Department of Health, the Arthritis Research Campaign, ZonMW and the Canadian Institute of Health Research). The framework is an input-process-output-outcome framework. It (typically) comprises two components: a definition of evaluation criteria (for outputs and outcomes of research) and a logic model.
UK Department for Innovation, Universities and Skills (DIUS)	UK	The "Economic Impacts of Investment in Research & Innovation" framework of the UK Department for Innovation, Universities and Skills (DIUS) aims to "assess the overall health of the science and innovation system, and how it delivers economic benefits".[60] It is the latest stage in a process of developing performance appraisal methods for the UK science and innovation system. The framework is used to model the delivery of economic impacts at the aggregate economy level through three stages and three

---

[60] DIUS (2007): "Economic Impacts of Investment in Research & Innovation July 2007"; downloadable from: http://www.berr.gov.uk/files/file40398.doc

| Frameworks	Country	Description
		influence factors.
European Union Framework Programme (EU)	EU	Framework Programme 7 of the European Union is meant as a key instrument contributing to the Lisbon, Gothenburg and Barcelona objectives – the system for evaluating the programme being a vector for tracking the results of research programmes and how they are contributing to the policy goals, and intended to be a way to identify what needs to be improved so that they can be more effective in achieving these goals.
		The responsibility for the evaluation of the Framework Programme rests with the evaluation unit in DG Research. It is supported by evaluation units in other DGs (JRC, INFSO, MARE, TREN, ENTR).
Congressionally Directed Medical Research Programs (CDMRP)	US	The Congressionally Directed Medical Research Programs (CDMRP) are part of the US Army Medical Research and Material Command (USAMRMC). The CDMRP manages (some of the) biomedical research that US Congress assigns to the USAMRMC.
		The CDMRP evaluation system consists of several elements. The three main ones are: its grants management system, its product database and its (breast cancer) Concept Award Survey.

Table 1 Evaluation Frameworks studied – Overview

Evaluation frameworks

In the following, a number of key elements of evaluation frameworks (arising from the frameworks studied) are discussed. First objectives, outcome measures, and level of aggregation of evaluation are examined. Subsequently, issues around timing and methodology are examined.

We suggest that these elements are highly interdependent. More specifically, we suggest that the choice of objective(s) (when establishing a research evaluation framework) influences the choice of outcome measures, and that the choice of outcome measures influences thinking about the right level of aggregation and timing. In addition, we propose that the level of aggregation influences the "choice of methods". For an illustration see (red lines in) figure below (Figure 2).
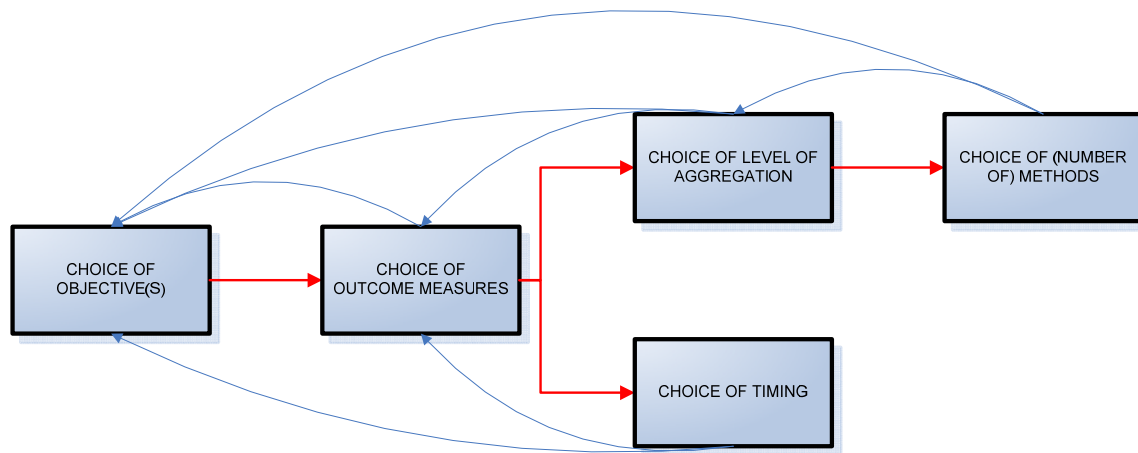
Figure 2 Outline of the argument

Each claim (with regard to the various relationships) is contrasted with a simple mapping of the frameworks studied. This should not be understood as a statistical test – because of the small sample size and because we do not control for other "explanatory" variables or "reverse" causality (illustrated by the various feedbacks in the figure above). Rather the arguments presented should be seen as propositions for further testing.

Objectives

The choice of an evaluation objective is of central importance. We suggest that many important decisions with respect to the development (and deployment) of a research evaluation framework are directly or indirectly influenced by the decision on what objective(s) to choose.

Earlier, four rationales for evaluation have been outlined: 1) to increase accountability (of researchers, policy-makers and funding bodies), 2) to "steer" the research process, 3) to provide a means for "advocacy", and 4) to provide an input into the management process (through better understanding and learning).

All four rationales have been picked up as "objectives" in the frameworks we studied. "Increased accountability" is stated as an objective in Buxton and Hanney (1996) for their Payback framework and for PART. "Steering" research is a central objective in the CDMRP framework. Advocacy is important in the Vinnova framework and the CDMRP framework. To use evaluation results as an "input" into the management process is stated as an objective by Buxton and Hanney for the Payback framework. It is stated also in the context of the LUMC framework, the framework of the European Union, DIUS, the CDMRP and Vinnova. An overview of the different frameworks and the corresponding objectives is given in the table below (Table 2).[61]

---

[61] It is important to note that the table lists only explicit objectives. For example, the fact that the PART framework uses "research targets" could be interpreted as implying an objective to "steer" research.

	Payback	DIUS	LUMC	MORIA	PART	Vinnova	EU	CDMRP
Increase accountability	✔				✔			
Provide "steering" of research process								✔
Provide input into the management process	✔	✔	✔			✔	✔	✔
Provide advocacy						✔		✔

Table 2 Evaluation Frameworks – Objectives chosen

No objective is listed for MORIA because it was designed for a different purpose (i.e. ex ante research evaluation) during peer-review evaluations of grant applications.

Output/outcome/impact measures

Once objectives are defined, measures upon which to base an evaluation need to be selected. The measures used in the evaluation frameworks studied can be categorized as follows:

- Input measures, capturing the resources consumed in the implementation of an intervention.
- Output measures, comprising the goods and services directly produced as a consequence of an intervention.
- Outcome measures, reflecting the initial impact of an intervention providing the reason for a programme.
- Impact measures, capturing the long-term changes an intervention brings about.[62]

---

[62] Please note that the terminology in the frameworks can differ from this definition. For the purpose of simplification, process measures are excluded.

	Payback	DIUS	LUMC	MORIA	PART	Vinnova	EU	CDMRP
Input measures	✓	(✓)						
Output measures	✓	✓	✓	✓	✓	(✓)		✓
Outcome measures	✓	✓	✓	✓	✓	(✓)	✓	✓
Impact measures	✓	✓		✓		✓	✓	

Table 3 Evaluation Frameworks – Outcome measures chosen

The table above (Table 3) gives an overview of measures used in each framework. It shows that only a few frameworks take into account the inputs going into the research process. (The brackets in case of DIUS indicate that inputs are measured but not linked to outputs, outcomes and impacts). Almost all frameworks measure outputs and outcomes. (The brackets in the case of Vinnova indicate that outputs and outcomes are relevant mainly at the monitoring and evaluation stage, not so much at the impact analysis stage). Impact measures are included in the DIUS and Vinnova frameworks (at macro level) and Payback and MORIA frameworks (at micro level).

For the purpose of simplification, we refer to: (i) outputs in combination with outcomes as upstream measures and (ii) outcomes in combination with impacts as downstream measures. Using "outcomes" both as part of upstream measures (when used in combination with "outputs") and as part of downstream measures (when used in combination with "impacts") seems to be justifiable since:

- In the former case (due to the focus also on "outputs") "outcomes" are likely to be more closely related to "outputs", whereas

- In the latter case (due to the focus also on "impacts") "outcomes" are likely to be more closely related to "impacts".[63]

The choice of outcome measures (i.e. whether upstream or downstream) is influenced, it can be argued, by what objectives have been chosen. More specifically, we suggest that the choice of an "accountability" and/or "advocacy" objective is likely to bias the choice of outcome measure towards more upstream measures (i.e. output/outcome measures) whereas the choice of a "steering" and/or

---

[63] Please note that there is no "double counting" of upstream measures and downstream measures. The reason is that an "outcome" is either counted as an upstream measure (if it is used in combination with outputs) or it is counted as a downstream measure (if it is used in combination with impacts). One way to think about this is by dividing outcomes into outcomes A-K which are associated more closely with outputs and outcomes L-Z which are more closely associated with impacts. If a framework uses outcomes in combination with both outputs and impacts, it is counted as "in between". See Payback framework below.

"learning" objective is likely to bias it towards more downstream measures (i.e. outcome/impact measures).

An accountability objective is likely to bias the choice of measures towards more upstream measures (i.e. outputs/outcomes) because downstream measures (i.e. outcomes/impacts) seem less appropriate in this context. One reason for this is that downstream effects often occur only 10–15 years after a research project has been completed – which can be too late for an evaluation with the aim to hold (for example) researchers accountable for their behaviour (since it may simply be too hard to track researchers down after such a long time).[64]

Another reason why downstream measures seem less suitable in the case of an accountability objective is that the long time lag between the end of a project and downstream effects (and, hence the many potential other influences which may have bearing on these effects) make it difficult to attribute a downstream measure to a certain researcher (funding body, or policy-maker). To the extent that a lower ability to attribute means a less adequate proxy for behaviour and, hence, a less adequate basis on which to hold people accountable, the choice of an accountability objective is likely to influence the choice of outcome measures (and biases it towards more upstream measures).

Similarly, an advocacy objective is likely to bias the choice of measures towards more upstream measures. The reason for this is, again, that downstream measures seem less appropriate in this context: 10–15 years after research has been completed (for downstream effects to occur) may be just too long to be useful (in terms of "signalling"). In addition (similarly to the case of accountability), to the extent that downstream measures mean a lower ability to attribute, and a lower ability to attribute means a less adequate proxy for behaviour and, hence, a less adequate basis to "signal" quality, the choice of an advocacy objective is (further) likely to bias the choice of outcome measures towards upstream measures.

A steering and/or learning objective, on the other hand, is likely to bias the choice of outcome measures towards more downstream measures. The reason for this is that "steering" and "learning" are likely to be driven by the variable of interest (and not so much by the variable which is (just) practical in terms of "holding accountable" or "providing advocacy").

The reason why policy-makers and research funders are likely to be interested to learn from, and to "steer" research towards downstream measures, is that they capture the downstream effects, which are what ultimately make a difference for people. (Upstream measures, on the other hand, are a less adequate proxy for these effects – because (for example) of the many unforeseeable contingencies influencing their development into downstream effects).

The figure below (Figure 3) supports this reasoning. It shows an association between accountability and advocacy objectives and upstream measures. It also shows an association between steering and learning objectives and downstream measures.

---

[64] Assuming that upstream measures are a less adequate proxy for downstream effects – (e.g.) because of the many unforeseeable contingencies influencing their development into downstream effects
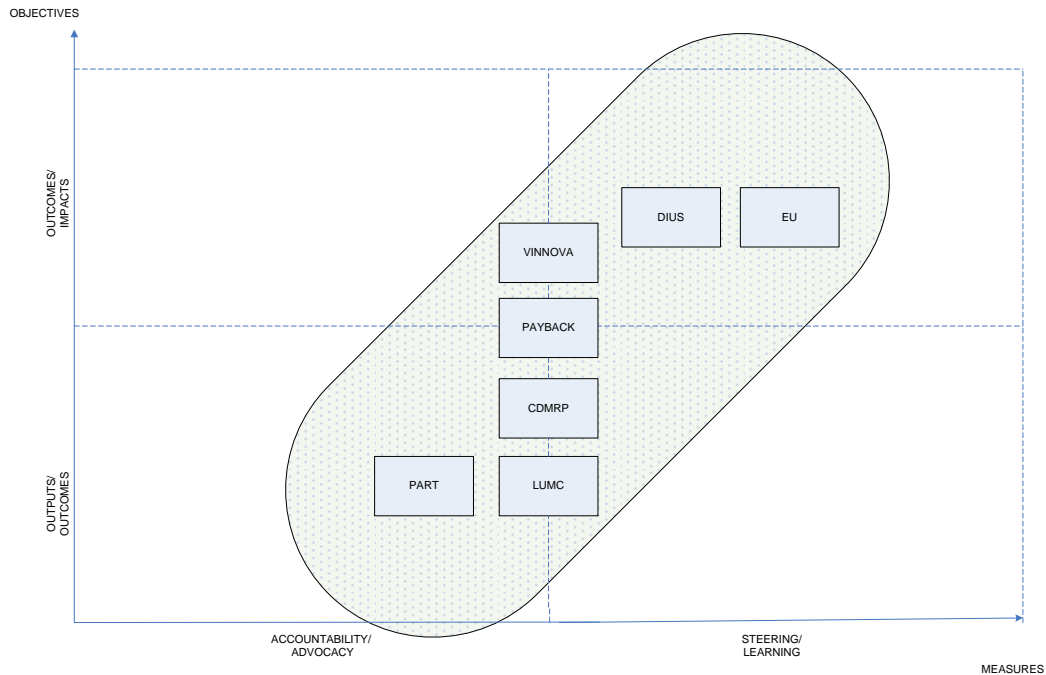
Figure 3 Evaluation Frameworks – by objectives and outcome measures

Categories of outputs, outcomes and impacts

Outcome measures (i.e. outputs, outcomes and impacts) can be categorized in different ways. This is typically done (using the phrasing of the LUMC framework) on the basis of "target groups" of research, comprising the research community, the general public, the public sector and the private sector. Correspondingly, research outputs, outcomes and impacts can be: scientific, social (including health-related effects), cultural and economic. See figure below (Figure 4) for an illustration.
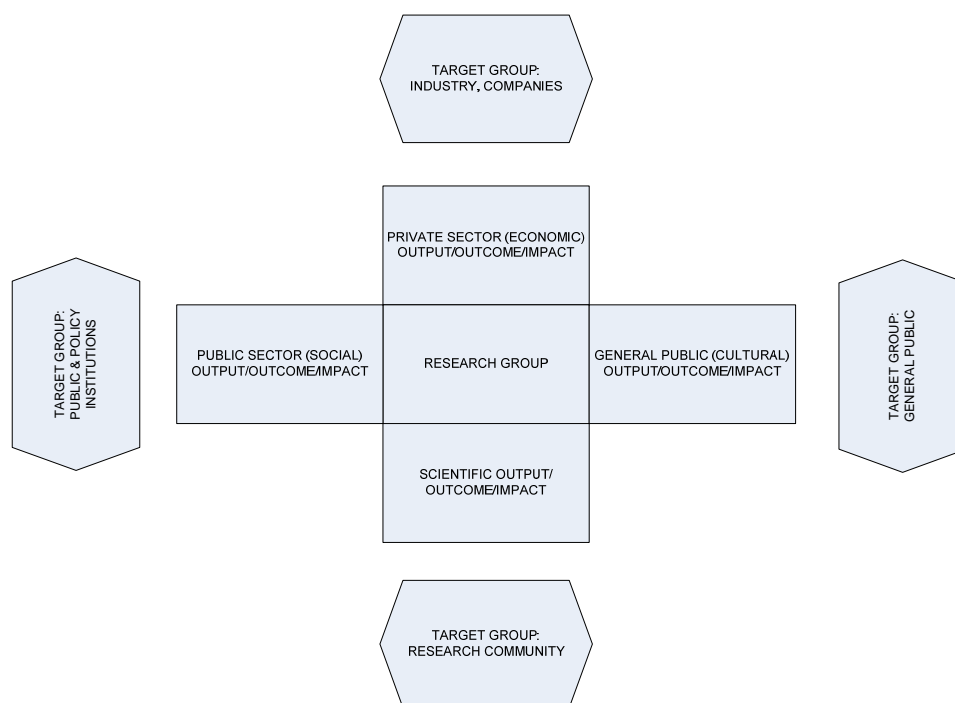
Figure 4 Target Groups adapted and modified from van Ark (2003)

The next figure (Figure 5) gives the frequency of the different categories in the frameworks. PART and the framework of the CDMRP do not group their outputs, outcomes and impacts and are, hence, not included in the figure.
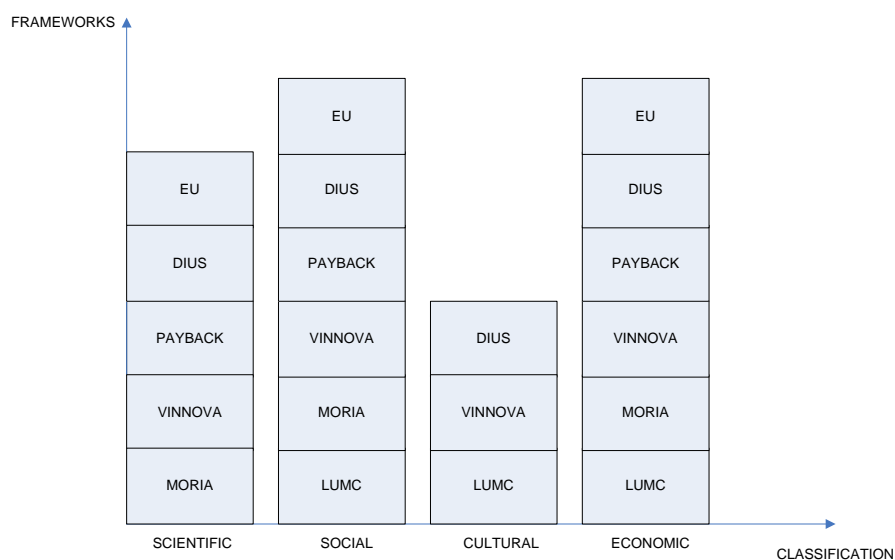
Figure 5 Evaluation Frameworks – Frequency of types of outcome

It is interesting to note that not only scientific outputs, outcomes and impacts are very popular in the frameworks studied, but also social and economic ones.

An explanation for this could be the combination of i) an increase in awareness of the importance of social and economic outputs, outcomes and impacts (of research) in the last decade or so[65] and ii) the insight that scientific measures of output, outcomes and impacts tell little about these "other" outputs, outcomes and impacts. As an example to illustrate the latter point: the fact that research on the cost-effectiveness of different incontinence pads is unlikely to be judged of high scientific impact tells us little about its social or economic benefits.[66]

Level of Aggregation

Having looked at the question "What to measure?", we can now look at "At what level to evaluate?". The level of aggregation in an evaluation can be low (individual researcher, research group or research project), intermediate (faculty or research programme) or high (research discipline, research council, charity, industry or university). An overview of the levels chosen in the frameworks studied is provided in the table below (Table 4).

	Payback	DIUS	LUMC	MORIA	PART	Vinnova	EU	CDMRP
High		✔				✔	✔	
Intermediate	✔				✔		✔	
Low	✔		✔	✔				✔

Table 4 Evaluation Frameworks – Level of Aggregation chosen

The table shows that all levels of aggregation are represented in the frameworks studied. The LUMC (research group), MORIA (researcher) and the CDMRP (project) evaluate at a low level of aggregation. PART (programme) and the EU framework (specific programme) choose an intermediate level for their evaluations. The Payback model has been applied both at a low level (grant) and intermediate level (programme). Vinnova (institute), DIUS (system) and the European Commission (Framework Programme) evaluate at a high level of aggregation.

It can be argued that the choice of outcome measures (itself influenced by the choice of objectives, as argued above) influences the choice of level of aggregation. More specifically, we suggest that downstream measures (i.e. outcome/impact measures) are likely to bias the choice of levels of aggregation towards higher levels, while upstream measures (i.e. output/outcome measures) are likely to bias it towards lower levels. The two cases are discussed in turn.

With regard to downstream measures: since (as argued above) downstream measures pose greater difficulty with regard to attributability, it is unlikely that they will be combined with low levels of aggregation – which also pose problems with regard to attribution. This is because an evaluator is

---

[65] Spaapen, J et al. (2007): *Evaluating Research in Context – A Method for Comprehensive Assessment*. The Hague, Consultative Committee of Sector Councils for Research and Development

[66] Smith, R. (2001): "Measuring the Social Impact of Research"; *BMJ*; 323; pp.528 ff.

unlikely to choose both an outcome measure that is difficult to attribute and a level of aggregation that makes attribution even more difficult.

Lower levels of aggregation are typically associated with more problems around attribution because of the "project fallacy": empirical evidence shows that a project often starts before the contracted work, continues after it, and integrates the contract work with a suite of other innovative activities which are funded elsewhere.[67] This suggests that the smaller the focus (or the lower the level of aggregation), the higher the chance that "other innovative activities" will be included (and falsely attributed) in an evaluation.

With regard to upstream measures (and the possible bias towards lower levels of aggregation), it seems that higher levels of aggregation are less compatible with upstream measures. Arnold et al find: "Evaluation does not get easier if we move from the project and programme level towards considering sub-systems and systems. The scale and complexity of the phenomenon mean that the same detail is not possible as when we operate at a smaller scale".[68]

This suggests that to the extent that studying upstream effects (occurring with relatively high frequency) is more detailed than looking at downstream effects (which are rarer and broader – not every output results in an outcome and/or impact), the choice of (upstream effects and consequently)[69] upstream measures is likely to bias the choice of levels of aggregation towards lower (less complex) levels.

The figure below (with the exception of MORIA) seems to confirm this reasoning. It shows an association of upstream measures with lower levels of aggregation and downstream measures with higher levels of aggregation.

---

[67] Georghiou, L. (2002): "Impact and Additionality"; in Boekholt, P. (2002): *Innovation Policy and Sustainable Development: Can Innovation Incentives make a Difference?*; IWT observatory.
[68] Arnold, E. et al. (2002): "Measuring 'relative effectiveness'"; in Boekholt, P. (2002): *Innovation Policy and Sustainable Development: Can Innovation Incentives make a Difference?*; IWT observatory.
[69] Assuming that upstream effects are best being measured by upstream measures.
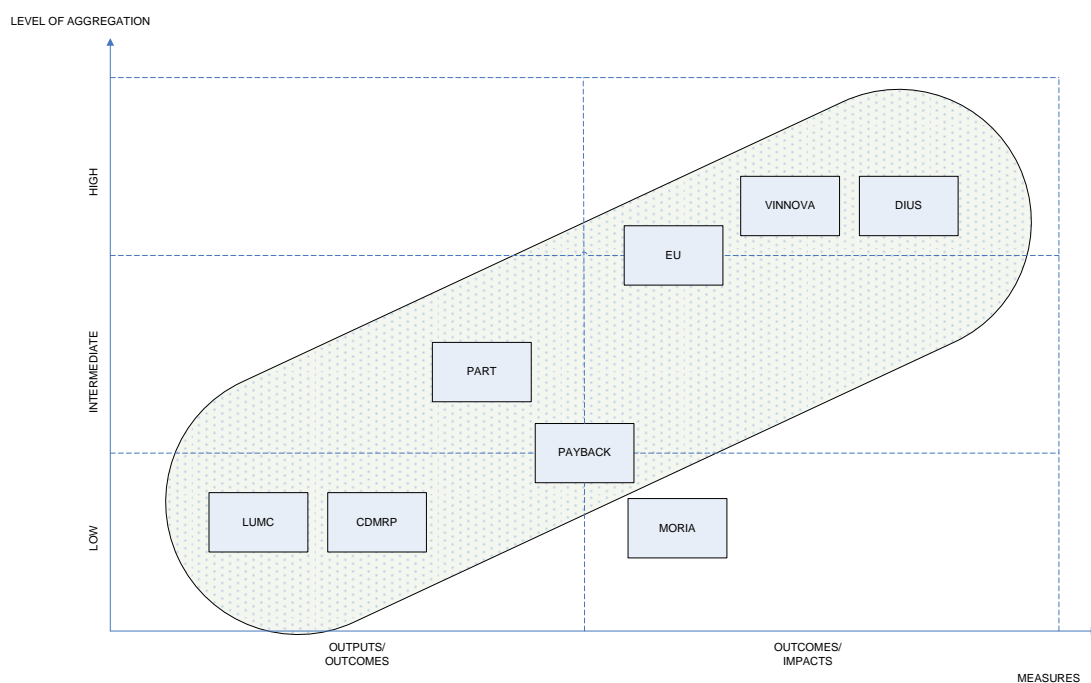
Figure 6 Evaluation Frameworks – by outcome measures and level of aggregation

**Timing**

Having discussed "What to measure?" and "Who or what to assess?", the next question is "How long after research is completed to measure/evaluate?". We have touched upon this question (and the trade-off with attribution) a few times already. Before going into this discussion, it is helpful to distinguish two ways of looking at evaluation related to timing.
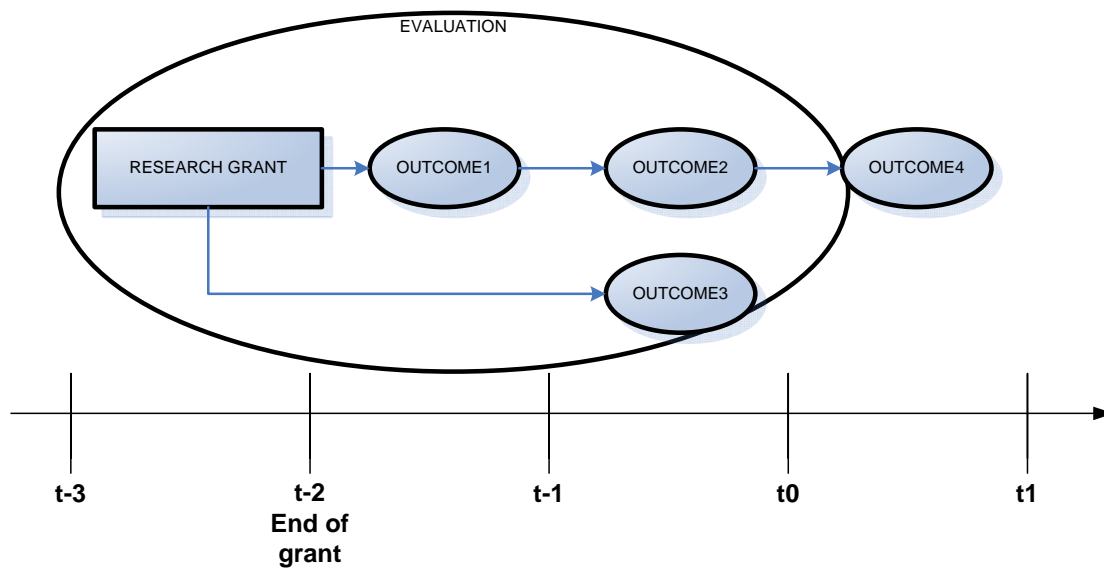
Figure 7 Longitudinal focus

The focus of an evaluation can be longitudinal or cross-sectional. That is, the evaluation can look at outputs, outcomes and impacts belonging to one piece (for example a project, programme or discipline) of research, or can be established within a certain time frame (for example by a group or institution) but not necessarily belonging to the same piece of research. The two concepts are depicted in the figure above (Figure 7 – Longitudinal focus) and below (Figure 8 - Cross-sectional focus). Note that "outcomes 1–4" in the figures can in fact be "outputs, "outcomes" or "impacts".


Figure 8 Cross-sectional focus

A-170
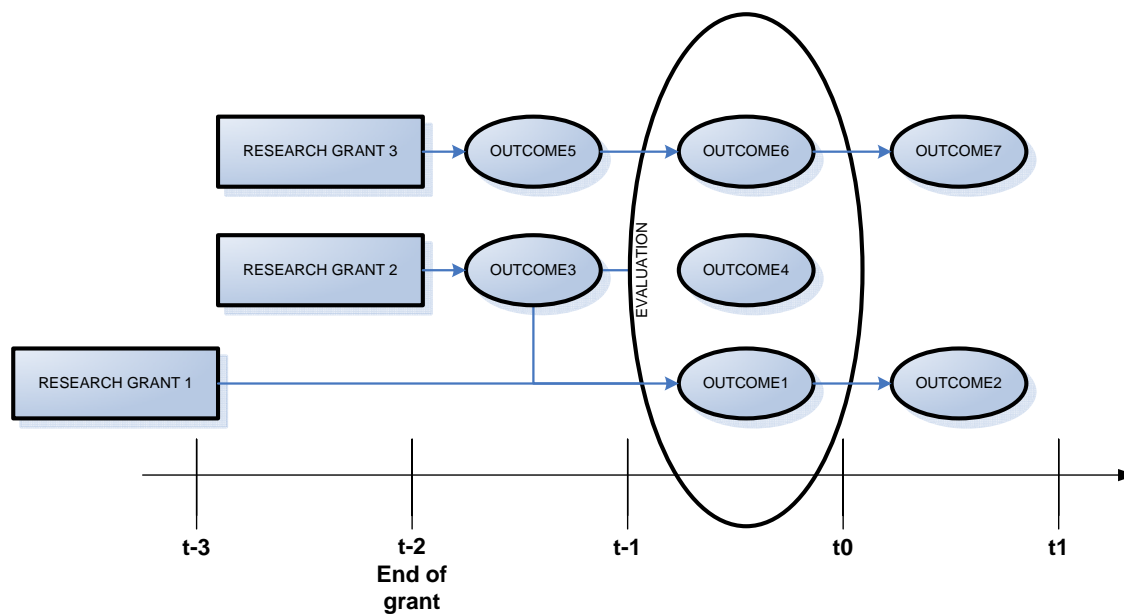
The two views (longitudinal and cross-sectional) are not mutually exclusive – but can coincide. This happens if the (cross-sectional) time span starts with the beginning of the longitudinal object of investigation, ends with the (longitudinal) evaluation period, and comprises the same individuals that are included in the object of study in the longitudinal evaluation.

We suggest that (regardless of whether the focus is longitudinal or cross-sectional) the timing of evaluation (i.e. the decision on how long after research to continue capturing outcomes) is influenced by the choice of outcome measures. The reason is that, typically, outputs, outcomes and impacts occur with different time lags after a project has finished. As an example, publications from specific research tend not to be published until a year or two after the project was finished. Patents for pharmaceutical products typically occur with a longer delay and the improvement in health (flowing from these products) often occurs only 20 years after the project was finished.[70]

The figure below (Figure 9, which plots upstream and downstream measures against timing) supports this reasoning. There is an association of upstream (i.e. output/outcome) measures with shorter evaluation time spans and of downstream (i.e. outcome/impact) measures with longer evaluation time spans.
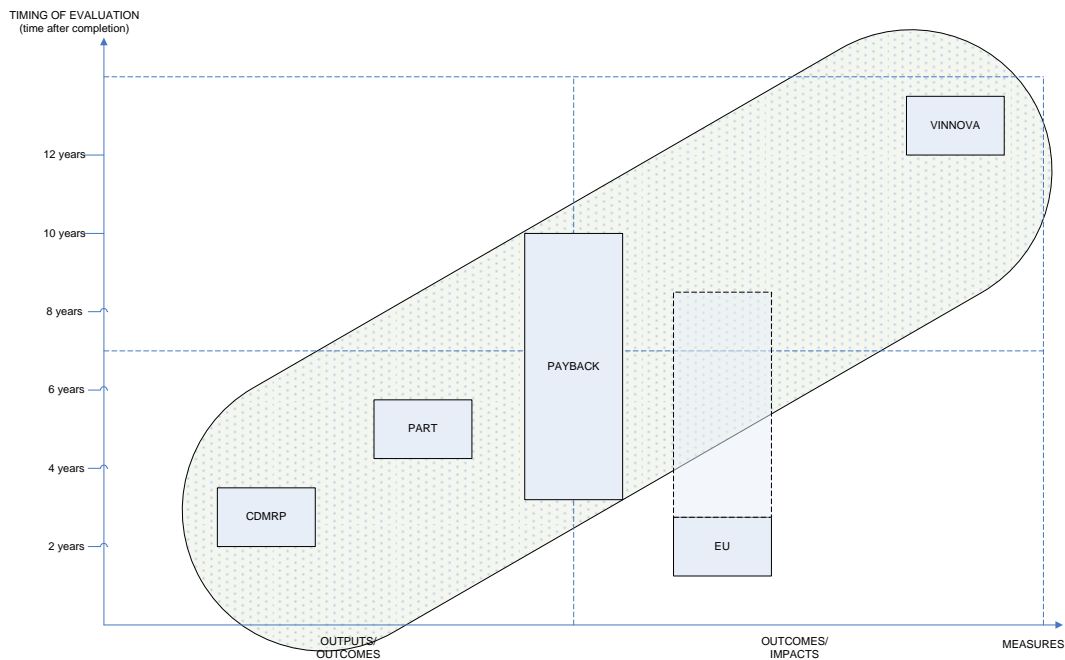
Figure 9 Evaluation Frameworks - by outcome measure and timing

---

[70] Braein, L. et al (2002): "The Norwegian systemic approach to impact estimation of R&D subsidies: focus on additionality and the contra-factual problem"; in Boekholt, P. (2002): *Innovation Policy and Sustainable Development: Can Innovation Incentives make a Difference?*; IWT observatory

MORIA, the LUMC framework and DIUS are not included in this figure. They all choose a cross-sectional (rather than longitudinal) focus. As a consequence, it is difficult to tell what their choice in terms of timing is. In longitudinal studies it is possible to infer "timing" from the choice of when to evaluate. This is not the case in evaluations with a cross-sectional focus (in which we can infer the time span used to search for outcomes – but not the span between research and evaluation).[71]

Timing considerations in evaluations based on the Payback model have varied across applications.[72] The timing of the EU framework (which is "no later than two years after a framework programme has been completed") is not perfectly consistent with the rest of the figure. One explanation (illustrated in the figure above) could be that the (present) framework programme spans seven years, which, with the two years after programme completion, amounts to a maximum of nine years between research and evaluation. This, it can be argued, makes it less important to have a long "waiting period" after programme completion.

How to measure

Having discussed issues around "What to measure?", "Who or what to assess", and "When to measure it?" we can now move on to the question "How to measure?". The table below (Table 5) gives an overview of the methods used in the frameworks studied.[73]

Following Fahrenkrog et al (2002), the rows of the table are divided into three parts: the first one summarizes methods around statistical data analysis, the second part comprises modelling methods, and the final part summarizes qualitative and semi-quantitative methods.[74]

All frameworks studied rely on at least one method summarized under semi-quantitative methods. Similarly, statistical data analysis methods are very popular in the frameworks studied. Modelling methodologies, on the other hand, are used (on a regular basis, at least) only in the DIUS and Vinnova frameworks and the framework of the European Union.

One possible explanation for the use of modelling techniques in the context of Vinnova, DIUS and the European Union, is the high level of aggregation (which these frameworks have in common). As mentioned before, the complexity of an analysis tends to increase with a higher level of aggregation, which, in turn, it can be argued, increases the need for more sophisticated methods.

The argument can be extended. That is, it can be argued that the level of aggregation not only influences how sophisticated the methods chosen are, but also how many different methods are used. The idea is that higher levels of complexity require more methods. Given that (i) a higher level of aggregation can

---

[71] It could be argued that timing (in the cross-sectional case) can be inferred from the start of (for example) a research group, but this seems unrealistic because of the problems of attribution this would entail, in particular for a long-established research group. Even if a group is not "long established", taking when it began as an indicator for "timing" is problematic. The reason is that such an approach implies a change in "timing" every year (which makes it hard to decide where, in the figure above, to place the respective frameworks).

[72] The study for the Arthritis Research Campaign (Wooding et al (2005): *Payback arising from research funding: evaluation of the Arthritis Research Campaign*), for example, covered 10–12 years after completion of research.

[73] The table should be seen as indicative (rather than affirmative), since some of the frameworks are in a (re-) development phase and may change the methods used (MORIA, LUMC, EU) or are by design very flexible as to which methods they rely on (DIUS, PART, Vinnova and EU).

[74] Fahrenkrog, G. et al (2002): *RTD Evaluation Tool Box – Assessing the Socio-Economic Impact of RTD – Policy*; IPTS Technical Report Series

be associated with a higher degree of complexity (as argued before) and that (ii) a higher degree of complexity can be associated with more methods, it is likely that the level of aggregation influences the number of methods used (and biases it towards higher numbers).

The reason why a higher degree of complexity is likely to require more methods is that this allows, as Polt et al. (2002) argue, to "fit" methods to particular dimensions of a problem (and hence to deal with it better). "The diversity of methodologies available for performing an evaluation are a signal of the multiple dimensions in which the impacts of policy intervention might manifest themselves. […]. Each methodology will be fitted to analyse particular dimensions of impacts, but the best evaluation approach would require a combination of various evaluation methodologies possibly applied at various levels of data aggregation".[75]

The figure below seems to support this reasoning. It shows that, on a higher level of aggregation (with more complexity) more methods are used than on lower levels of aggregation (with arguably less complexity). Of course the list of methods is not comprehensive and could have been structured in ways that would have influenced the mapping. Nonetheless, the result seems interesting – if only as an indicative one.
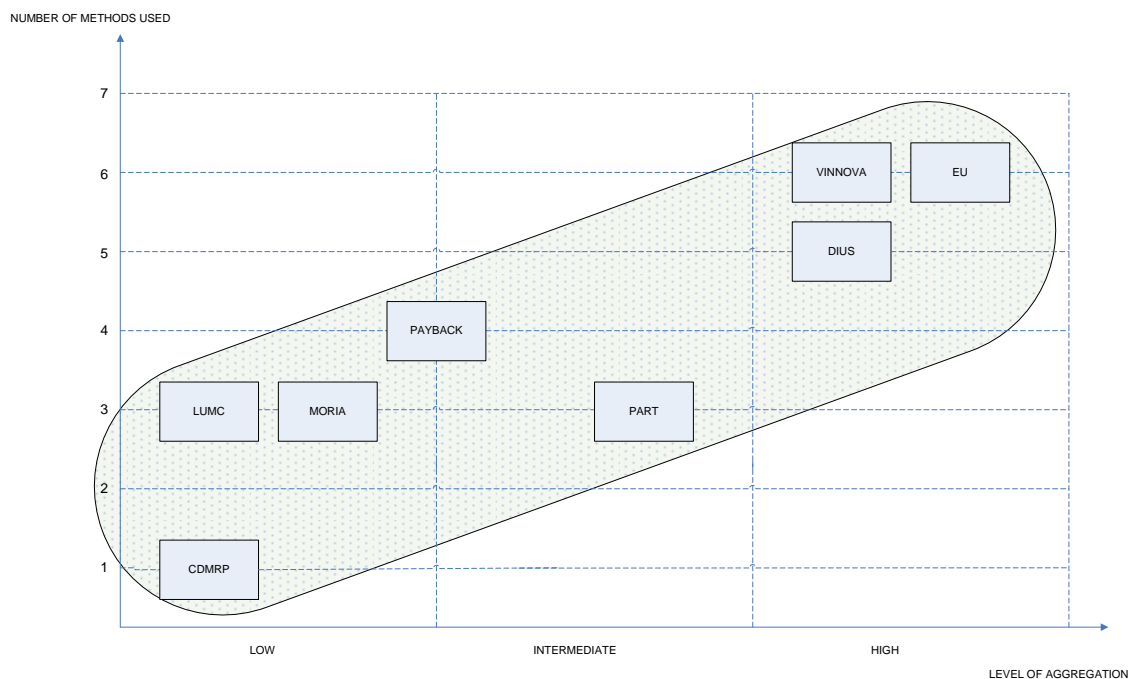
Figure 10 Evaluation Frameworks – by level of aggregation and number of methods used

---

[75] Polt, W. et al (2002): "The purpose of Evaluation"; in Fahrenkrog, G. et al (2002): *RTD Evaluation Tool Box – Assessing the Socio-Economic Impact of RTD – Policy*; IPTS Technical Report Series p.72

Methodologies	Brief Description	Payback	DIUS	LUMC	MORIA	PART	Vin-nova	EU	CDMRP
Statistical data analysis									
– Questionnaire	– provides basic data to describe the research process, outputs, outcomes and impacts	✔	✔	✔	✔	✔		✔	
– Benchmarking	– allows performance of comparisons based on a relevant set of indicators		✔	✔	✔	✔			
Modelling methodologies									
– Macroeconomic modelling	– allows estimation of broader socio-economic impacts of policy interventions		✔				✔	✔	
– Microeconometric modelling	– allows estimation of outputs, outcomes and impacts at the level of the individual						✔		
– Productivity analysis	– permits assessment of the impact of R&D on productivity growth at different levels of data aggregation.		✔				✔		
– Control group approaches	– allows capture of the effect of a project, programme or policy on participants using statistical sophisticated techniques.								
Qualitative and semi-quantitative methodologies									
– Interviews and case studies	– uses direct observation of events to investigate behaviours in their indigenous social setting.	✔	✔				✔	✔	✔
– Cost-benefit analysis	– allows establishment of whether a policy, programme or project is economically efficient by appraising all its economic and social effects.							✔	
– Expert Panels/Peer Review	– measures scientific output, outcome and impact relying on the perception scientists have.					✔	✔	✔	
– Bibliometrics (and other quant. indicators)	– allows measurement of scientific output and outcome, drawing on information on publications (patents, research funding etc.).	✔		✔	✔		✔		
– Network Analysis	– allows analysis of the structure of cooperation relationships and the consequences for individuals' decisions.								
– Logic modelling	– used to capture the logical flow between inputs, outputs, outcomes and impacts	✔							
– Foresight/Technology Assessment	– used to identify potential mismatches in the strategic efficacy of project, programmes and/or policies.								

Table 5 Evaluation Frameworks – Methods used – similar to Polt et al. (2002)

Conclusion

In this (first part of the) report we identified five key elements of research evaluation: evaluation objectives, outcome measures, levels of aggregation, timing and evaluation methods. We found significant differences along these key elements between the evaluation frameworks we studied.

In addition, we suggested (and provided some evidence in this direction) that these elements are not independent from each other - but that trade-offs exist when choosing them. An important conclusion following from this is that these key elements ought to be chosen very carefully – taking into account that elements which appear appropriate in isolation need not constitute a good choice in combination with other key elements.

In particular, the choice of an evaluation objective is important. We suggested that it, directly or indirectly, influences the appropriateness of all other key elements. More specifically, we suggested that the choice of an evaluation objective influences the choice of outcome measures, and that the choice of outcome measures influences thinking about the right level of aggregation and timing. In addition, we proposed that the level of aggregation influences the "choice of methods".

Each claim was contrasted with a mapping of the eight evaluation frameworks we studied. The mappings (by and large) supported our reasoning. It is important to note, however, that this is no conclusive evidence (in any statistical sense) but only a starting point for further research.

A note on Additionality

An interesting finding from the frameworks studied is the absence of the question of additionality in most cases. It has long been realized that what an evaluation asks needs to go beyond the level of effects achieved by the beneficiaries of a policy (such as researchers) and pursue the issue of what difference (relative to no intervention) is made by that policy (programme, project etc.).[76]

Conceptually, additionality appears relatively simple on superficial examination. It involves comparison with the counterfactual – what would have happened if no intervention had taken place. Georghiou (2002) has developed a more fine-grained picture. He differentiates between:

- Input additionality, which is concerned with, for example, whether for every euro provided in support, at least an additional euro is spent on the target activity (i.e. on research – as opposed to higher salaries, for example)

- Output/Outcome additionality, which, is concerned with the proportion of outputs/outcomes that would not have been achieved without support

- Behavioural additionality, which looks at how research support changes the way in which a project is carried out (for example, how it influences the pursuit of new areas of enquiry in research activity).[77]

---

[76] Georghiou, L. (2002): "Impact and Additionality"; in Boekholt, P. (2002): *Innovation Policy and Sustainable Development: Can Innovation Incentives make a Difference?*; IWT observatory.

[77] "The UK Department of Trade and Industry has articulated these changes in three sub-divisions – scale additionality when the activity is larger than it would otherwise have been as a result of government support (perhaps creating economies of scale); scope additionality, where the coverage of an activity is expanded to a wider range of applications or markets than would have been possible without government assistance (including the case of creating a collaboration in place of a single company effort); and acceleration additionality when the activity is significantly brought forward in time, perhaps to meet a market window." Georghiou, L. (2002): "Impact and Additionality"; in Boekholt, P. (2002): *Innovation Policy and Sustainable Development: Can Innovation Incentives make a Difference?*; IWT observatory.

Output/Outcome additionality has been touched upon in the Payback model (using a quasi-experimental design)[78] and the framework of the EU (asking programme participants directly about the counterfactual). The EU framework also addresses the issue of behavioural additionality (by means of its questionnaire). The Vinnova framework discusses both forms of additionality. All other frameworks are, by and large, tacit about the issue.

One possible way to think about additionality in the context of this report is illustrated below. The idea is that the choice of a type of additionality may (to some extent) be influenced by the choice of outcome measures (i.e. output, outcome or impact).
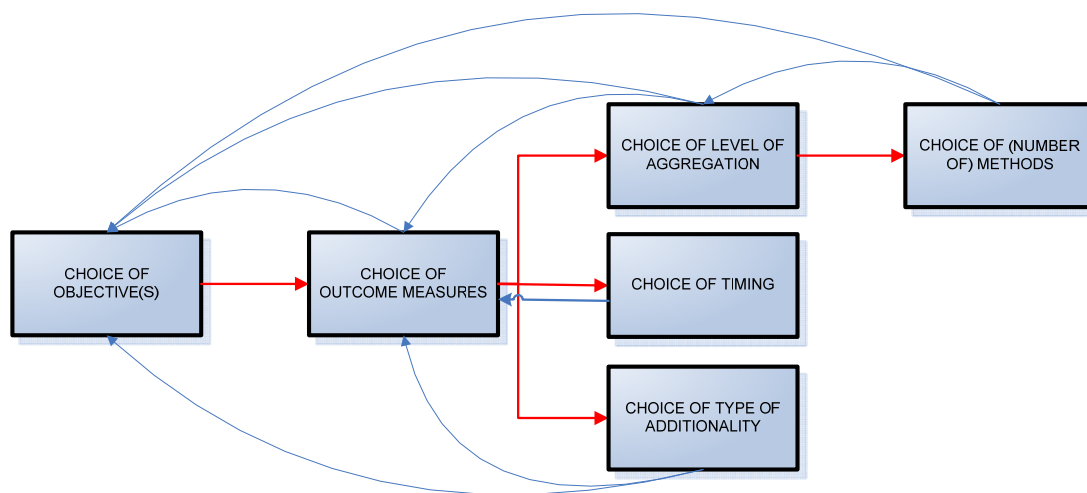
Figure 11 Including additionality in the discussion

One reason why the choice of outcome measures could influence the choice of a type of additionality is that a focus on downstream measures seems to be in conflict with that on behavioural additionality. In fact, Hervik found a trade-off between economic impact and behavioural additionality (in a study of successive policies in Norway).[79] A possible reason for this, suggested by Georghiou, is that "high [behavioural] additionality may easily be associated with an increased risk […] because the intervention has tempted a [researcher, research group etc.] to move beyond its competences or to undertake a project which was more risky than usual" (and, hence, having a lower impact).[80]

Since the trade-off between impacts and behavioural additionality need not imply anything with regard to the relationship between upstream measures and behavioural additionality (not having an impact does not mean that there cannot be an output, even an outcome), behavioural additionality may well be consistent with frameworks choosing output/outcome measures (and not impact measures). Because of the absence of a discussion of additionality, the frameworks examined do not allow for this question to be addressed further at present. This could be a starting point for future research.

---

[78] Wooding et al (2005): *Payback arising from research funding: evaluation of the Arthritis Research Campaign*
[79] Hervek, A. (1997): "Evaluation of user-oriented research in Norway: the estimation of long-run economic impacts in Papaconstantinou" in Polt, W. et al (1997): *Policy Evaluation in Innovation and Technology – Towards Best Practices*, OECD
[80] Georghiou, L. (2002): "Impact and Additionality"; in Boekholt, P. (2002): *Innovation Policy and Sustainable Development: Can Innovation Incentives make a Difference?*; IWT observatory

Case Studies

LUMC:

1. Introduction

The framework in place at the Leiden University Medical Center (LUMC) is an ex post evaluation framework which focuses on the "societal impact" of research at the level of the research group. Looking at "societal impact" (rather than scientific quality), the framework can be seen as part of a broader movement in the Netherlands to correct for the "serious imbalance in the research portfolio" (arising from a sole focus of evaluation on scientific quality).

The underlying assumption of the framework is that societal impact and scientific quality need not always go hand in hand. Smith explains: "Quality to scientists tends to mean originality of subject, thought, and method. Much research that scientists judge of high quality has no measurable impact on health – often because the lag between the research and any impact may be decades. Thus scientists would think of the original work on apoptosis (programmed cell death) as high quality, but 30 years after it was discovered there has been no measurable impact on health. In contrast, research that is unlikely to be judged as high quality by scientists – say, on the cost-effectiveness of different incontinence pads – may have immediate and important social benefits."[81]

2. Basic Description

The first thing to note about the LUMC framework is that it is concerned only with the evaluation of "societal impact". Scientific quality is assessed in a different exercise carried out by the Centre for Science and Technology Assessment (CWTS). (A study by Mejer and Mostert (2007) shows that a comparison of the results from the two exercises can bear interesting findings.)

Drawing on the work by van Ark and Klasen, the basic idea of the framework is to understand evaluation of research outcomes as "valuation of communication of the research group with its surroundings" – where "valuation of communication" focuses on three *modes of communication*:[82]

       1) knowledge products,

       2) knowledge exchange & esteem, and

       3) knowledge use.

and the *surroundings* comprise:

       1) public sector,

       2) private sector, and

       3) the general public.

---

[81] Smith, R. (2001): "Measuring the Social Impact of Research"; *BMJ*; 323; pp.528 ff.
[82] Van Ark, G. (2007): *Societal Impact Evaluation of Research Groups: The Communication Metaphor*; Presentation at the Sigtuna Workshop on Economic Returns of Medical Research Nov. 2007.

The evaluation is based on indicators, which can be structured (as in the table below (Table 6)) along "modes of communication" (columns) and "surroundings" (rows).

	Knowledge products	Knowledge exchange & esteem	Knowledge use	Attractiveness
Public sector (also social impact)	+prof. publications +guidelines +procedures etc.	+ prof. input in R&D +prof. functions +prizes +lectures etc.	+prof. citations +prof. use of guidelines, etc.	Revenues generated (from prof. training, courses and R&D contributions etc.)
Private sector (also economic impact)	+patents +knowledge products and services	+formal co − operations +lectures and courses for companies etc.	+use & sale of patents, +products & services	Revenues generated (from contract research, private research contributions etc.)
General public (also cultural impact)	+lay publications +media attention etc.	+public input in R&D, public functions +prizes etc.	+public citation of publications +use & sale of knowledge products & services etc.	Revenues generated (from charity funding, public R&D contribution etc.)

Table 6 LUMC "Modes of Communication" and "Surroundings"

It is important that the evaluation goes beyond the mere categorization of indicators. A scoring system is used to translate a research group's performance for each indicator in a (standardized) numerical score. This allows comparison and aggregation of indicators across different modes of communications and surroundings.

For example, it allows the comparison of the "value" of communication of a research group with the public sector ("social impact") by means of knowledge products with the communication of the group with the private sector ("economic impact") flowing from knowledge products, or the comparison of the "value" of communication of the group with the general public ("cultural impact") through knowledge exchange and esteem with that flowing from knowledge use. In addition, the scoring system allows the production of an overall score for the "value" of communication of the group across all modes of communication and surroundings.

The "value" of communication refers to the societal impact of research. The different indicators are weighted accordingly (i.e. on the basis of their expected translation into societal impact). This means, for example, that a publication in a local newspaper gets a lower score (in the system) than one in a national one, since it has a lower reach and hence, most probably, lower impact.

"Attractiveness" is listed as a separate column in the table above. It is not meant to be a separate "mode of communication", however. Instead, it is a category to capture indicators that are considered particularly important (and, hence, should get a high weighting factor). More specifically, the column summarizes the revenues generated from research outputs (in the context of all modes of communication). This is considered particularly important since it reflects a high interest in and, hence, high impact of research.

The weighting of different indicators is not only based on the expected translation of certain outputs into societal impact/use but also takes into account the relative scarcity (in terms of occurrence) of

certain outputs. For example, two indicators, which *a priori* would be considered of equal importance with regard to their expected translation into societal impact, may end up with different weighting factors if performance with regard to one is generally much lower than with regard to the other.

3. Background

The LUMC framework builds upon the (theoretical) work of The Royal Netherlands Academy of Arts and Sciences[83], Gerrit van Ark's work,[84] and the work of the Health Council (Dutch Department of Health).[85]

It was commissioned by Professor Klasen, Dean of the LUMC, in 2006. It was developed (for the LUMC) by Gerrit van Ark the same year. Its implementation started in 2007 and was led by Ruud Kukenheim and Stéfan Ellenbroek (LUMC Directorate of Research). They received support from Prof Klasen and Gerrit van Ark as well as from Ingeborg Meijer and Bastian Mostert from Technopolis.

Currently, evaluation is on "active" pause. The reason for this is problems with the electronic data collection system. It is hoped that the framework will be adopted at other medical centres in the Netherlands which would allow the sharing of development costs for a new, better data collection system as well as benchmarking (of the different medical centres). A (further) likely development of the framework concerns the indicators in use. At the moment the framework comprises 98 (sub-) indicators, which is felt to be too many. It seems likely that a reduction in the number of indicators will occur in the near future.

4. Technical aspects

Objectives
The central objective of the framework is to inform policy-makers on the societal usefulness of research. As discussed earlier, this can be interpreted as providing input into the management process as well as a way to demonstrate that policy objectives are met.

Attribution
As indicated in part I of the report (despite the use of the term societal *impact*), the indicators used are rather "upstream" (i.e. closer to "output" than "impact", as defined earlier). This reduces the problems of attribution, insofar as upstream measures tend to occur earlier (than downstream measures) and, hence, tend to be affected by fewer factors other than the one of interest.

The fact that the framework looks at research groups – independently from research grants – can also help to avoid problems of attribution (since outputs do not have to be linked to specific (potentially sequential) grants). At the same time, a potential problem may arise if individual researchers or even research groups move from one medical centre to another (since then their output might be attributed to the new centre, despite the fact that most of the efforts have been undertaken at the old one.

Costs
The development costs for an electronic data collection system are expected to be around €100K. The costs of running the system are expected to be around half a day of work per department per year – which adds up to 20 days for the whole centre per year; adding 3–4 days for central processing and analysis this gives 23–24 days in total per year.

It is hoped that the development costs for the ICT system can be shared between different medical centres. The actual evaluation costs fall on each centre.

---

[83] "Societal Impact of Applied Health Research"

[84] Van Ark, G. (2007): *Societal impact of R&D*; Den Haag, ZonMw

[85] Dutch Health Council (2007): "Research that matters. Responsiveness of University Medical Centers to Issues in Population Health and Health care" downloadable from: http://www.gr.nl/samenvatting.php?ID=1651

Consequences of the evaluation
The findings from the framework are used to inform (together with findings from the evaluations of scientific quality) the management process concerned with the future strategy of the LUMC. The findings are not meant, however, to provide a basis for hard and fast rules to make strategy (and funding) decisions.

Stakeholder involvement
Evaluatees (i.e. the ones being evaluated) provide input into the evaluation framework. They are also involved in the development of the framework through representatives on the "Scientific Board" (a body which, among other things, discusses (potential) issues arising from the evaluation). Finally, evaluatees' experiences from the pilot studies have been taken into consideration in the development process of the framework.

MORIA:

1. Introduction

MORIA stands for "measure of research impact and achievement". It looks at outputs, outcomes and impacts of research across three domains: "knowledge", "health gain" and "economic benefits". MORIA was developed at the Australian NHMRC as an analytic (support) instrument in the (ex ante) peer review process for grant applications. It builds on the Record of Research Achievement (RORA) framework. At the moment, it seems unlikely that MORIA will be used in this (ex ante evaluation) function. Some of the work may, however, be used in the NHMRC post grant assessment.

A particularly interesting aspect of MORIA is its scoring system. Similar to the LUMC framework, findings are translated into a (standardized) numerical score. This allows comparison and aggregation of findings across projects and (within projects) across different domains.
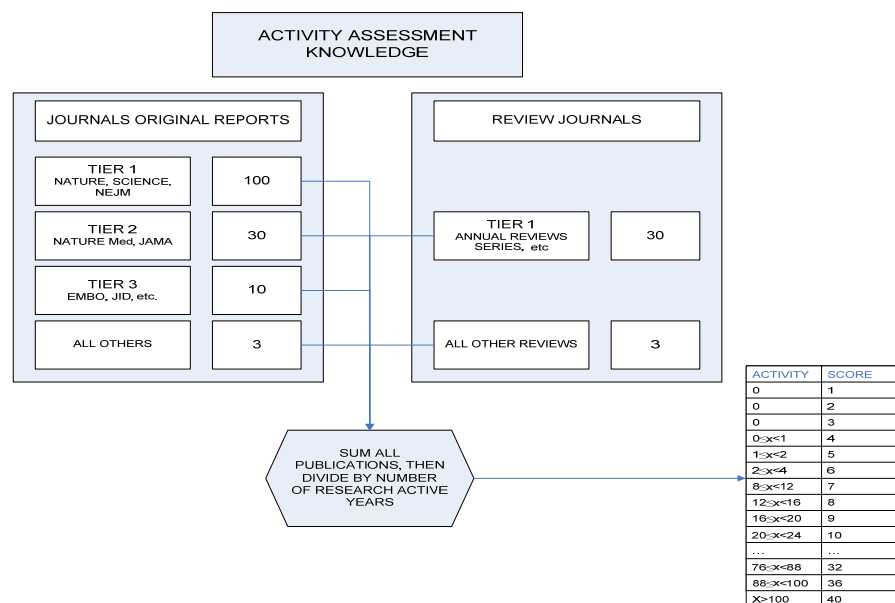
2. Basic Description

MORIA looks at outputs ("activity"), outcomes ("recognition") and impacts of research across three domains: "knowledge", "health gain" and "economic benefits", as illustrated in the table below (Table 7).

		Domain		
Level	**Score**	**Knowledge contribution**	**Health gain**	**Economic benefit**
Activity	1–40	+ Publication counts weighted by journal rankings etc.	+ Health sector engagement	+ Patents, industry engagement etc.
Recognition	8–150	+ Count of highly cited publications etc.	+ Recognition in clinical and public health practice	+ income, savings, employment
Impact	100–200	+ Up to 3 substantial impacts on knowledge	+ Up to 3 substantial impacts on health	+ Up to 3 substantial commercial achievements

Table 7 MORIA – Overview

For each cell, an assessment is conducted. The figure below (Figure 12) shows how this is done in the case of outputs (or "activity") in the context of knowledge contribution. The idea is to count publications, weight them according to journal ranking, and then divide the resulting score by the number of research active years (which, in a further step, can be translated into an "activity score").[86]

---

[86] The fact that scores are divided by research active years reflects the fact that MORIA was designed as an ex ante evaluation framework taking a "whole of career approach" to assess the track record of a researcher.

ACTIVITY ASSESSMENT
KNOWLEDGE

JOURNALS ORIGINAL REPORTS

TIER 1 NATURE, SCIENCE, NEJM	100
TIER 2 NATURE Med, JAMA	30
TIER 3 EMBO, JID, etc.	10
ALL OTHERS	3

REVIEW JOURNALS

| TIER 1 ANNUAL REVIEWS SERIES, etc | 30 |
| ALL OTHER REVIEWS | 3 |

SUM ALL PUBLICATIONS, THEN DIVIDE BY NUMBER OF RESEARCH ACTIVE YEARS

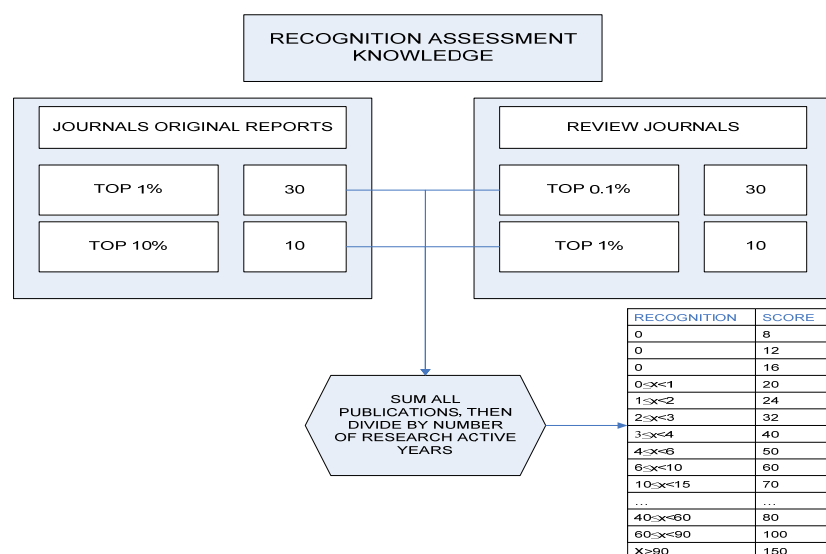| ACTIVITY	SCORE
0	1
0	2
0	3
0≤x<1	4
1≤x<2	5
2≤x<4	6
8≤x<12	7
12≤x<16	8
16≤x<20	9
20≤x<24	10
…	…
76≤x<88	32
88≤x<100	36
X>100	40

Figure 12 MORIA – Activity Assessment

The activity assessment of "health gain" follows the same logic. The only difference is that, rather than looking at publications, "engagement" (i.e. direct involvement in politics and practice as a result of research) and "translation products" (such as policy documents and clinical guidelines) are considered.

A citation analysis is used to assess the outcome (or "recognition") of knowledge generated. Points are allocated (depending on the relative performance with regard to citations) and a "recognition score" calculated (taking into account the number of research active years). See figure below for an illustration.

It is important to note that the recognition score is based on field-adjusted performance in citation centiles. (In particular, the ISI 104 field list was found to provide much better results than the ISI 24 field list). Another option that was discussed was that each article for an individual could be assigned to a field based on ISI's field designation for that journal – this would reduce applicant "gaming", but add to the complexity in terms of analysis.

RECOGNITION ASSESSMENT KNOWLEDGE

JOURNALS ORIGINAL REPORTS

| TOP 1% | 30 |
| TOP 10% | 10 |

REVIEW JOURNALS

| TOP 0.1% | 30 |
| TOP 1% | 10 |

SUM ALL PUBLICATIONS, THEN DIVIDE BY NUMBER OF RESEARCH ACTIVE YEARS

RECOGNITION	SCORE
0	8
0	12
0	16
$0 \leq x < 1$	20
$1 \leq x < 2$	24
$2 \leq x < 3$	32
$3 \leq x < 4$	40
$4 \leq x < 6$	50
$6 \leq x < 10$	60
$10 \leq x < 15$	70
…	…
$40 \leq x < 60$	80
$60 \leq x < 90$	100
$X > 90$	150

Figure 13 MORIA – Recognition Assessment

The assessment of recognition with respect to "health gain" follows a similar pattern. Rather than looking at citations, however, "adoption" performance (internationally, nationally and locally) is considered.

The "impact" assessment process with respect to "knowledge contribution" is depicted below. The basic idea is to allow researchers to make a case for their work (i.e. to what extent it is of "broadest and deepest significance"). On the basis of this "case", an "impact score" is allocated (with a higher weight given to research of "global importance" rather than "field-specific importance"). The assessment of health impacts follows the same ("make a case") logic.

IMPACT ASSESSMENT KNOWLEDGE

LIST UP TO 3 DISCOVERIES THAT YOU WOULD LIKE CONSIDERED IN THE CATEGORY OF IMPACT.

PROVIDE EVIDENCE THAT YOU PLAYED A CENTRAL AND CRUCIAL ROLE IN MAKING THE DISCOVERIES AND THAT THE DISCOVERIES ARE OF THE BROADEST AND DEEPEST SIGNIFICANCE.

KNOWLEDGE IMPACT

| GLOBAL IMPORTANCE | 100 |
| MAJOR FIELD-SPECIFIC IMPORTANCE | 30 |

SUM ALL IMPACTS, THEN DIVIDE BY NUMBER OF RESEARCH ACTIVE YEARS

IMPACT	SCORE
30	100
$30 \leq x < 100$	150
$X > 100$	200

Figure 14 MORIA – Impact Assessment Knowledge

The assessment of economic benefits is described in the table below (Table 8). "Activity" is assessed on the basis of indictors (such as number of patents, number and size of consultancy work, and other

contract work). "Recognition" draws on (among other things) commercial income, investment and employment data.

Economic Benefits			
Activity	+ patents	+ Commercial/Economic engagement (measured as income from consultancy work, research contracts etc)	+ Discovery development (measured as government grants to aid commercialization, industry collaboration etc.)
Recognition	+ Commercial Income or Health Savings	+ Capital Investment	+ Employment
Impact	+ List up to 3 commercial/economic achievements for assessment of Impact + Evidence must include verifiable external criteria, e.g. long-term viability of a major company, major profits, income received, long-term savings on the health budget.		

Table 8 MORIA – Economic Benefits

3. Background

The development of MORIA started in late 2003 with the establishment of a small working group by the NHMRC Research Committee. The aim of the group was to develop a standardized approach to measure the "track record" of NHMRC funding schemes. The group comprised researchers from basic science, clinical, public health and health services research disciplines and bibliometric expertise.

Where available, the working group has taken into account relevant international publications and an analysis of current NHMRC funding scheme criteria to develop the new metric. The work by Jonathan Grant and colleagues at RAND Europe, initially for the Arthritis Research Council in the UK, informed the work of the group. The development of the Australian Government's Research Quality Framework (RQF), and its focus not just on research quality but also on impact, was significantly informed by the NHMRC thinking arising from the MORIA development.

In August 2004, a workshop for researchers across a range of disciplines was held to provide comments on the results of the group. On the basis of this, the working group further refined the MORIA prototype. Since late 2007 MORIA has been on hold. It is unlikely that it will be used in the peer review process for grant applications (as a measure for researchers' "track record"). There is, however, the possibility that some of the work on MORIA will be used within the NHMRC to develop an evaluation framework for post-grant research outcomes.

4. Technical aspects

Objectives
MORIA's stated objective is to produce a reliable measure of research impact and achievement that is logically feasible and transparent. It was never intended to be used on its own during the peer review process but only to aid and assist the NHMRC peer review process (to make it more efficient and effective).

Pilot studies
There have been pilots in the different domains (knowledge, health gain, economic benefit): A pilot study of the "knowledge" domain was conducted in late 2005 and early 2006, with a sample of 30 individuals currently in receipt of NHMRC grant funding but only for the basic science area. The sample

was chosen to represent a range of seniority and experience in applicants to the various grant-funding vehicles. The pilot data showed that the "activity" and "recognition" scores of the knowledge component were relatively easy to assign in the basic science area. Moreover, the pilots showed that the scoring system displays good discrimination between applicants with differing levels of output, and does not appear to be adversely affected by the age of the applicant. A comparison of the activity and recognition scores with the citation per publication rate of the individuals in the pilot test revealed no strong relationship. This indicates that the MORIA activity and recognition scores were not simply reproducing information that could be derived from bibliometrics.

A pilot test of the "economic benefit" domain was undertaken in mid-2006 to determine the facility of such a model. A sample of 20 NHMRC applicants with known commercial research experience was chosen, and 11 responses provided. The pilot only collected data on the activity"and recognition levels. Results from the pilot test indicated that the approach taken thus far is feasible. There may be (smaller) issues around confidentiality, the dollar values assigned to each of the levels in the recognition area (to provide better discrimination between outputs), and scaling (in order to avoid clustering of respondents at the top and bottom ends of the scales).

The "health gain" has not had any pilot testing to date. There was, however, a group identified to develop further the entire health gain domain of MORIA. This stalled with the rest of the programme in late 2007.

Data collection
The collection of much of the data for the knowledge component of MORIA relies on the citation databases provided by the Institute of Scientific Information – which is part of Thomson Scientific. The citation data from publications has recently (also) become accessible through the Endnote Web (another part of Thomson Scientific). Endnote Web is a web-based bibliographic and research tool that allows an end user to collect and compile a citation library online. End-note could allow an applicant to provide information on his or her track record – saving a great deal of workload normally placed on the NHMRC grant reviewers. Most other aspects rely on self-reporting (with externally verifiable evidence).

Costs
Since MORIA is not in regular use, there is no cost data available. No cost estimates have been done, to our knowledge.

Stakeholder involvement & feedback
Evaluatees provide input into the evaluation framework. In addition, to the extent that MORIA is meant to be part of a larger peer review process (which typically allows for various feedbacks), evaluatees are involved in the overall process as well.

The feedback from NHMRC's Research Committee was largely positive with respect to the general principles of MORIA. There was a good deal of concern over the use of MORIA to develop a numeric score for grant applications, as this was seen as a threat to the subjective nature of current peer review mechanisms. There was also a good deal of concern around the ability of MORIA to be extended beyond basic science grants to public health and clinical medicine grants, as it was suggested that the outcomes of these areas were sufficiently different from what was expected from basic science, and that, hence, MORIA would need major redevelopment for these applications.

PART:

1. Introduction

PART stands for Program Assessment Rating Tool. It was introduced shortly after George W. Bush took office in 2001, as part of his agenda to improve government management. PART is used to assess the effectiveness of around 800 federal programmes. It takes the form of a diagnostic questionnaire.

An interesting element of PART is that it evaluates programmes (to a large extent) on the basis of performance goals. To do so it adopts output, outcome and efficiency measures. Most weight is on outcome measures. The idea is that "Outcome measures are most informative, because these are the ultimate results of a program that benefit the public. Programs must try to translate existing measures that focus on outputs into outcome measures by focusing on the ultimate goals of a program […]."[87] Yet, an exception is made for research and development programmes. The OMB guidance finds that outcome measures may be inappropriate in this context, since "results [often] cannot be predicted in advance of the research".[88]

2. Basic Description

PART (at the NIH and in general) takes the form of a diagnostic questionnaire used to rate selected programmes. The questionnaire contains 25–30 general questions about each of the following four broad topics to which all programmes are subjected:

- Programme purpose and design (20%): to assess whether the programme design and purpose are clear and defensible. (Sample questions: Does the programme address a specific and existing problem, interest or need? Is the programme designed so that it is not redundant or duplicative of any other federal, state, local or private effort?)

- Strategic planning (10%): to assess whether the agency sets valid annual milestones and long-term goals for the programme. (Sample questions: Does the programme address a specific and existing problem, interest or need? Is the programme designed so that it is not redundant or duplicative of any other federal, state, local or private effort?)

- Programme management (20%): to rate agency management of the programme, including financial oversight and programme improvement efforts. (Sample questions: Does the programme use strong financial management practices? Does the programme collaborate and coordinate effectively with related programmes?)

- Programme results (50%): to rate programme performance on goals reviewed in the strategic planning section and through other evaluations. (Sample questions: Has the programme demonstrated adequate progress in achieving its long–term performance goals? Does the programme demonstrate improved efficiencies or cost-effectiveness in achieving programme goals each year?)

Each section carries a (pre-specified) weight (see above) resulting in a total weighted numerical rating ranging from 0 to 100. In addition, programme managers can alter weights within each category to emphasize key factors of the programme. To avoid manipulation of the total score, weights must be adjusted prior to responding to any question. Based upon the numerical scores, OMB assigns a management and performance rating to the programmes. These range from the highest rating of "effective", to "moderately effective", to "adequate", to a lowest score of "ineffective". In addition, the

---

[87] Gilmour, J.B. et al (2006): "Assessing performance assessment for budgeting: The influence of politics, performance, and program size"; *Journal of Public Administration Research and Theory*.
[88] Ibid p.72

rating of "results not demonstrated" means that the measures developed were not adequate to determine the programme's effectiveness.

Suggested answers to the questions (along with explanations and evidence) are provided by programme officials. A budget examiner for the programme then reviews the materials submitted, and decides which answers to give for each of the questions. Federal agencies (such as the NIH) have the opportunity to formally appeal the answers with which they disagree. Appeals are considered and adjudicated by a five-person panel comprised of members of the President's Management Council, a group of deputy secretaries responsible for management issues at their respective agencies. As an example, the table below (Table 9) gives the recent PART assessments of NIH programmes.

PART Year	Year Conducted	Programme	Score	Rating	Summary
FY 05	FY 03	HIV/AIDS Research	83	Moderately Effective	The HIV/AIDS Research Program was deemed *moderately effective*. Improvements based on PART included a scientific update to the deadline for the end target, and an increase in the number of programme evaluations submitted for the planning and budget development process.
FY 06	FY 04	Extramural Research	89	Effective	The Extramural Research Program was deemed *effective*. The PART resulted in integrating the CJ and GPRA Plans/Reports and led to discussions addressing budget performance alignment. Programme exemplifies good design, planning, management and results.
FY 07	FY 05	Intramural Research	90	Effective	The Intramural Program was deemed *effective*. Programme exemplifies good design, planning, management and results.
FY 07	FY 05	Building & Facilities	96	Effective	The Building and Facilities Program was deemed *effective*. Building and Facilities received the highest numerical score. There were no programme flaws noted.
FY 08	FY 06	Research Training	N/A	Effective	The Research Training Program was deemed *effective*. Programme is effective at training and retaining researchers in the biomedical research field.
FY 08	FY 06	Extramural Construction	N/A	Moderately Effective	The Extramural Research Facilities Construction Program was deemed *moderately effective*. Programme effectively manages construction and renovation projects from the pre-award phase and during construction.

Table 9 PART – NIH Programme Assessment

3. Background

Shortly after George W. Bush took office in 2001, he committed to an agenda of improved government management. A key element of this agenda was to make the government more results-oriented by expanding the use of performance budgeting. He directed the Office of Management and Budget (OMB) to work with each agency to recast its budget to include performance information. In 2003, he expanded this effort by committing to a programme-by-programme assessment of performance. He directed the OMB to lead this assessment effort (as well). In response, the OMB developed an assessment

framework, with the assistance of agencies and outside experts, which it named the Program Assessment Rating Tool, or PART.

In February 2006, OMB unveiled a new website, www.ExpectMore.gov, that makes available the assessments of all programmes that have been subjected to PART. ExpectMore.gov divides programmes into two groups: those that are "performing" and those that are "not performing". By exposing programmes that are not performing, OMB hopes to compel them to improve, and to give their constituents and stakeholders arguments to demand improvements. These efforts have been recognized by the broader government improvement community. In 2005, PART was awarded a Ford Foundation Innovations in American Government award.

PART builds upon the Government Performance Results Act (by using the supply of performance information that federal agencies have been generating as a result of GPRA).[89] Yet, PART goes beyond GPRA in two important ways. Firstly, PART renders judgement on whether programmes are effective. Secondly, PART enables decision-makers to attach budgetary and management consequences to those programmes that cannot demonstrate their effectiveness.

4. Technical aspects

Objectives

PART has two main objectives. The first one is to provide decision-makers with the information they need to allocate scarce resources in a way that will yield the greatest benefit. The second objective is to induce organizational change. That is, to encourage agencies to find better ways of achieving their goals and improving their results. A further objective (often linked to the second goal) is for PART to introduce a new level of transparency. OMB's new website, www.Expect More.gov, in which it makes available the assessments of about 800 programmes that have been subjected to PART, can be seen as a step in this direction.

Attribution

As mentioned before, PART puts a lot of emphasis on "outcome" measures. The benefit of this is that it focuses attention towards the "ultimate goal of a program". At the same time, "outcomes" are typically further removed from what programmes directly influence (and may have causes other than the programme) and so an attribution problem may occur.

The programmes are assessed and reassessed on a five-year schedule. PART acknowledges that in some cases this may be too short for results to be reflected in "outcome" measures. Possible ways to deal with this problem (within PART) are to use output measures and/or "measures towards an outcome".

Consequences of the evaluation

One aspect of the consequences of the assessment is manifested in PART's improvement plan: up to three PART follow-up actions are included in each programme assessment summary. The improvement plan is developed in collaboration between the OMB and the federal agencies.

---

[89] As for the GPRA framework, the NIH collects information in five functional areas: 1) scientific research outcomes, 2) communication and transfer of results, 3) capacity building and research resources, 4) strategic management of human capital and 5) programme oversight and improvement.

In each area it sets strategic goals (typically for 6 years). These are selected according to (different) criteria. In case of the scientific research outcomes (1) that is, representativeness, meaningfulness, specificity, objectivity and reportability.

In addition, an important goal of PART is to link budget decisions with assessments of outcomes and overall programme quality. At the same time, it is important to note that a number of factors contribute to a programme's budget request, and so the assessment score in and of itself does not determine funding recommendations.

Stakeholder involvement & feedback

Evaluatees are involved at several stages of the process: They provide suggested answers and evidence for the questionnaire. As described above, evaluatees have also the possibility to appeal the assessment. In addition, if evaluatees can demonstrate significant improvement, they can request a reassessment to improve the rating of their programme.

Gilmour finds that PART is taken very seriously at the programme and bureau level. "Management systems imposed from above always meet a certain amount of scepticism and resistance, and that is true with PART. But attitudes have changed as programme managers have seen the determination and persistence of OMB in implementing PART. […] the analysts and programme managers interviewed by the author – virtually all careerists – almost uniformly believed that the exercise of completing the PART questionnaire was good for programmes."[90]

---

[90] Gilmour (2007): "Implementing OMB's Program Assessment Rating Tool (PART): Meeting the Challenges of Integrating Budget and Performance"; downloadable from: www.businessofgovernment.org/pdfs/GilmourReport.pdf p.30
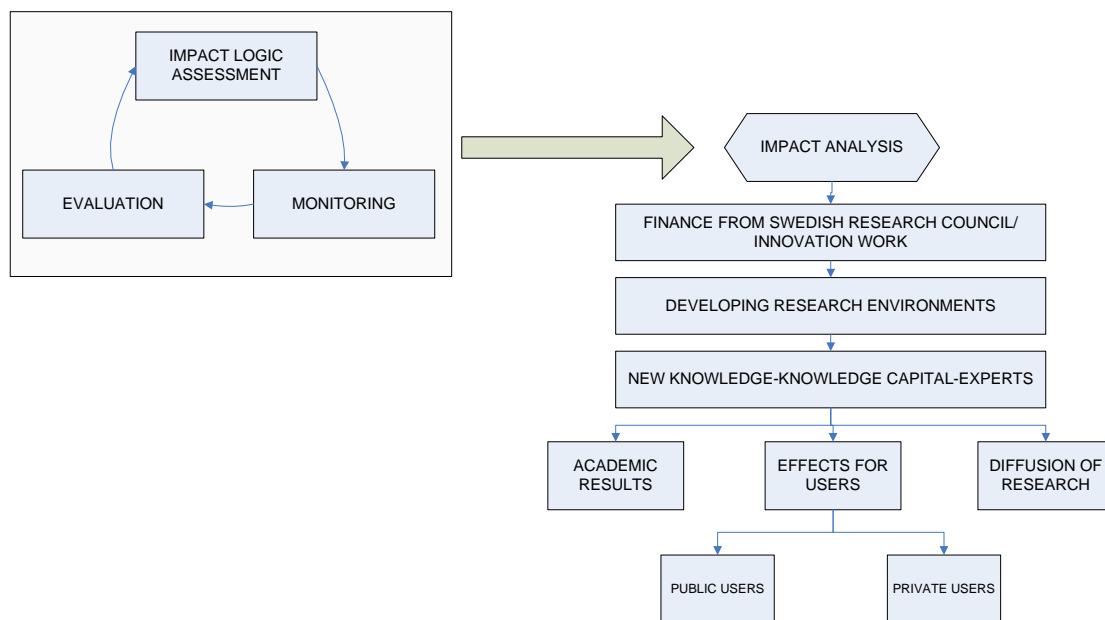
Vinnova*:*

1. Introduction

Vinnova is the Swedish governmental agency for innovation systems. When Vinnova was formed in 2001, there was an interest in understanding better what its initiatives are achieving, as well as in developing methods to estimate its long-term impacts. Since 2003 Vinnova has been conducting impact analyses on a yearly basis to respond to this interest.

The Vinnova framework consists of two main parts: an ongoing evaluation process and an impact analysis. There is some variation in how the framework is applied. The discussion in this report is based on the very recent work on traffic safety.

2. Basic Description

The two main parts of the Vinnova framework are depicted in the figure below (Figure 15), with the ongoing evaluation process in the upper left-hand corner.

Figure 15 Vinnova – Overview

The idea underlying the "ongoing evaluation process" stage is to define the results and impacts of a programme against which it can be evaluated, and define corresponding indicators. In addition, it allows the collection of data which can later be used in the impact analysis.

The ongoing evaluation process comprises three parts: an impact logic assessment, monitoring, and evaluation of the project. The "impact logic assessment" is an ex ante assessment. Its main purpose is to ensure that the initiative in question can be evaluated and that the evaluation generates policy-relevant information.[91] The "monitoring" provides continuously updated information on the development of a programme. In addition it provides early indicators of impacts. "Evaluation" concentrates on clarifying whether the goals for a programme are being or have been achieved. The results of the evaluation are used as the basis for deciding on changes to ongoing programmes or as a starting point for the design of new programmes. Moreover, their findings feed into the impact analysis.

---

[91] "A well-implemented impact logic assessment leads to conclusions on which information needs to be gathered during the course of a programme as well as what the main evaluation issues will be in various evaluations." Vinnova's focus on impact.

Impact analyses form the core of the Vinnova framework. They are conducted to study the long-term impact of programmes (typically a whole portfolio). The right-hand side of the figure above shows the main channels through which impacts (are assumed to) manifest themselves: academic results, public users, private users and diffusion of research.

More specifically, impact through "academic results" considers:
- If the content has "answered society's needs" (evaluated through a panel of experts)
- If research is at a high academic level (looking at impact factors[92] and PhD supervision (assuming that the latter indicates the success in transferring acquired expertise to the next generation)).
- If researchers actively participate internationally (looking at, among other things, the number of grants from the EU Framework Programme for research going to Swedish researchers, and participation in ISO-committees [assuming that this helps to spread research results]).

Impact through "public users" looks at the effect of research when put into practice through politics. The impact can be estimated in four steps:
- In a first step, data on the actual development of an issue (e.g. traffic safety) is collected and plotted.
- In a next step, on the basis of previous research, impacts of various factors (on traffic safety) are collected.[93]
- The findings from the second step can then be used to plot a "counterfactual" development (such as the development of traffic safety in the absence of (some or all of) the impacts considered).
- In a third step, finally, the two developments (actual and "counterfactual") can be compared (to get an idea of the (combined) impact of the measures on traffic safety).

The idea is illustrated in the figure below (Figure 16).
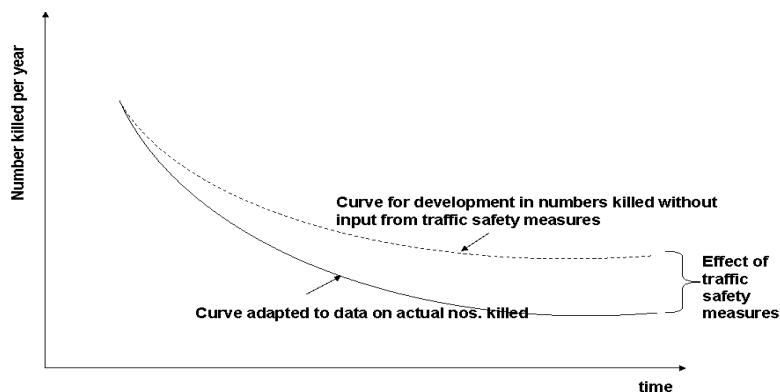
Figure 16 Vinnova – Impact through public users

---

[92] The ISI impact factor is a measure of how many times an average article in a given publication in the large ISI research database is cited in journals in the database in the course of a given year. The annual impact factor is the relationship between citations and the number of published articles. The impact factor is calculated by dividing the number of citations in a given year by the number of citable units published in the two preceding years.

[93] Interaction between the measures is not considered.

Impact through "private users" considers:
- The ratio between costs and benefits for consumers, companies and society as a whole (where the business economic profit at the national level is part of the calculation). This can be done on the basis of case studies (using "willingness-to-pay" data to get economic units).
- The possible industry-related benefits of increased exports. This can be measured/proxied by using production (and installation) costs (assuming benefits from exports are at least as great as costs).

Impact through the "diffusion of research in society" looks at:
- How research influences (national) thinking (assessed on the basis of case studies) and
- How it influences policy-making (looking at how often the research is referenced in policy documents, or to be found on governmental websites).

3. Background

Vinnova's predecessors used monitoring and evaluation, but paid little attention to long-term impacts. When Vinnova was formed in 2001, there was an interest in better understanding what its initiatives were achieving, as well as in developing methods to estimate its long-term impacts. In autumn 2001 four pilot-type impact analyses were conducted,[94] the main purpose of which was to develop various methods for future analyses. The pilot studies were carried out by Technopolis Ltd, Vinnova and Goran Friborg.

Since then, Vinnova has produced seven impact analysis reports. The impact analyses differ in significant ways. This is due to learning (some studies had the explicit "subsidiary aim" to develop and test new methodologies) as well as differences in the areas studied. Since 2003, in response to the requirement of the Swedish Ministry of Enterprise, impact analyses have been conducted on a yearly basis.

4. Technical aspects

Objectives
Vinnova's ultimate goal is to "promote sustainable growth through funding of need-driven research and development of effective innovation systems". The aim of its impact analyses is to demonstrate its success in achieving this goal – in a way that is transparent and "understandable" to non-experts in the field.

Data collection
Much of the data collection occurs during the monitoring process. This is done by Vinnova's programme managers. Their search is typically informed by pilot projects. This involves using a few projects to get an idea of what information needs to be gathered, how this (gathering process) can best be organized, and what indicators work for particular cases. Other sources include: interviews, group discussion, documents and literature, as well as data collected (originally) for different purposes.

Costs

The costs for an impact analysis (including data gathering) lie between €150K and €200K

Stakeholder involvement

---

[94] See Technopolis Ltd., Friborg, G. and Vinnova (2002): *Impact of Vinnova's predecessors' support for needs-driven research. Four impact analyses during the period 1975–2000*; Vinnova Innovation in Focus VF2002:1, "Foreword", p.1

Researchers are involved in impact analyses at an early stage, to help identify key channels of impacts, and to help identify (expected) impacts. In addition, after the impact analyses are completed, the results are (typically) discussed in workshops comprising researchers, policy-makers and other stakeholders.

Payback*:*

1. Introduction

The Payback framework was developed at the Health Economic Research Group at Brunel University (HERG). It has been applied in a number of different contexts and with different research funders (including the UK Department of Health, the Arthritis Research Campaign, ZonMW and the Canadian Institute of Health Research).

The framework is an input-process-output-outcome framework. It (typically) comprises two components: a definition of evaluation criteria (for the outputs and outcomes of research) and a logic model.

2. Basic Description

The two components of the framework are: a definition of evaluation categories for the outputs and outcomes of research, and a logic model of the research process.

A categorization of Payback is illustrated in the table below (Table 10). It comprises knowledge, research benefits, political and administrative benefits, health sector benefits and broader economic benefits.

A. Knowledge
B. Benefits to future research and research use:
i. Better targeting of future research; ii. Development of research skills, personnel and overall research capacity; iii. Critical capability to utilize appropriately existing research, including that from overseas; iv. Staff development and educational benefits.
C. Political and administrative benefits:
i. Improved information bases on which to take political and executive decisions; ii. Other political benefits from undertaking research.
D. Health sector benefits:
i. Cost reduction in the delivery of existing services; ii. Qualitative improvements in the process of service delivery; iii. Increased effectiveness of services, eg increased health; iv. Equity, eg improved allocation of resources at an area level, better targeting and accessibility; v. Revenues gained from intellectual property rights.
E. Broader economic benefits:
i. Wider economic benefits from commercial exploitation of innovations arising from R&D; ii. Economic benefits from a healthy workforce and reduction in working days lost.

Table 10 Payback – Categorization

The framework makes extensive use of indicators to assess each of these categories. A list of exemplary measures for each category is provided in the table below (Table 11).

A. Knowledge

       i. Number of publications resulting from research

       ii. Peer review rankings of results of funded research.

       iii. Bibliometric measures

B. C. Political and administrative benefits:

       i. Number of public policies influenced

       ii. Number of practice guidelines

       iii. Number of products receiving regulatory approval after sponsored trails.

D. Health and health sector benefits:

       i. Public health: Strategic research initiatives and their outcomes.

       ii. Quality Adjusted Life Years (QALYs)

       iii. Cost savings in the provision of health care

       iv. Patient satisfaction

E. Broader economic benefits:

       i. Commercialization: Number and nature of patents, spin-off companies and licences for intellectual property generated from funded research; Income from IP commercialization.

       ii. Direct cost savings: Estimates of the value of high-impact innovations developed through research.

       iii. Human capital: Reduction in productivity loss through illness or injury due to innovations from research.

Table 11 Payback – Exemplary Measures

The second component of the Payback framework (i.e. the logic model) consists of nine steps (seven stages and two interfaces) as shown below (Figure 17). Its purpose is to indicate how, and at what stages, the Payback categories can be assessed: usually "knowledge" production and "benefits to future research" are associated with stage III ("primary outputs"), "political and administrative benefits" with stage IV ("secondary outputs"), "health and health sector benefits" as well as "broader economic benefits" with stage VI ("final outcomes"). It is important to note that this reflects broad correlations (rather than a perfect match). Similarly, the (high degree of) linearity underlying the (logic) model is meant to give an indication of the different assessment stages (and not so much to specify an exact research translation process).
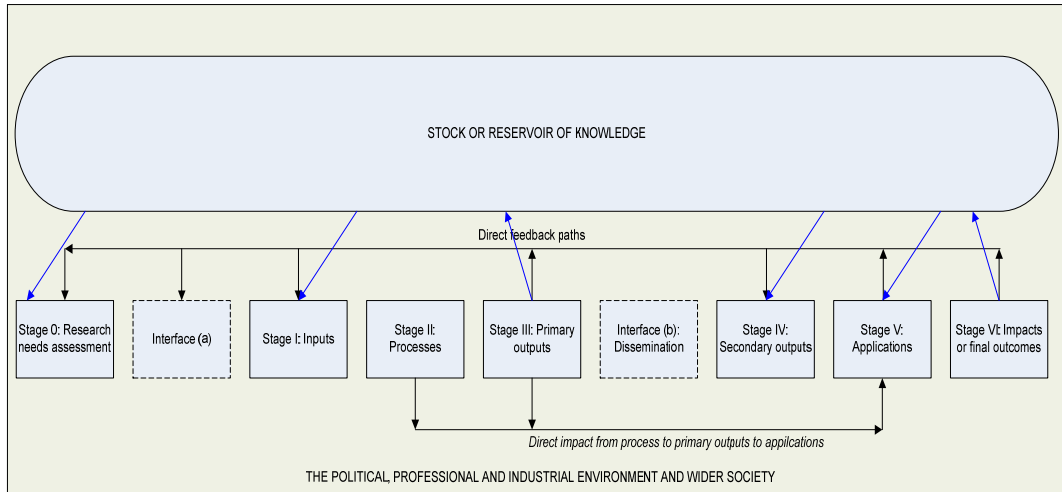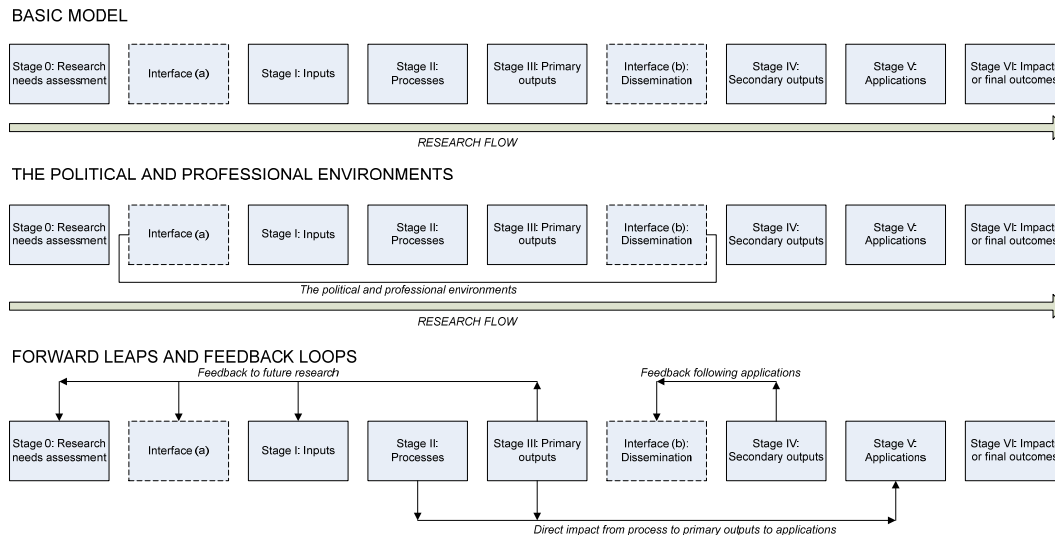
Figure 17 Payback – Logic Model

3. Background

The Payback was originally commissioned by the UK Department of Health in 1993 to evaluate the health service research that it supported. Subsequently the Payback framework has gone through a number of iterations and applications. The first phase of the work, described in Buxton and Hanney (1994, 1996) and Buxton et al (1994), consisted of:

- a categorization of Payback under five headings and

- a nine-stage model – as above, as well as

- eight case studies to test this categorization and modelling.

The second phase of the study confirmed that the multidimensional categorization of Payback, as originally presented under the five headings listed above, was (by and large) robust. Similarly, in reviewing a further 10 case studies, it was shown that the nine-step model was valid, but the issue of whether the scientific endeavour can be modelled as a linear process and the importance of the political and professional environment were raised. This led to further refinement of the Payback model as illustrated below (Figure 18). From this basis, the Payback framework has been applied in a number of different contexts, extended and developed further by HERG and RAND Europe.

BASIC MODEL

| Stage 0: Research needs assessment | Interface (a) | Stage I: Inputs | Stage II: Processes | Stage III: Primary outputs | Interface (b): Dissemination | Stage IV: Secondary outputs | Stage V: Applications | Stage VI: Impacts or final outcomes |

RESEARCH FLOW

THE POLITICAL AND PROFESSIONAL ENVIRONMENTS

| Stage 0: Research needs assessment | Interface (a) | Stage I: Inputs | Stage II: Processes | Stage III: Primary outputs | Interface (b): Dissemination | Stage IV: Secondary outputs | Stage V: Applications | Stage VI: Impacts or final outcomes |

The political and professional environments

RESEARCH FLOW

FORWARD LEAPS AND FEEDBACK LOOPS

Feedback to future research   *Feedback following applications*

| Stage 0: Research needs assessment | Interface (a) | Stage I: Inputs | Stage II: Processes | Stage III: Primary outputs | Interface (b): Dissemination | Stage IV: Secondary outputs | Stage V: Applications | Stage VI: Impacts or final outcomes |

Direct impact from process to primary outputs to applications

Figure 18 Payback – Refinement

The framework has been used on an ad hoc basis by (among others) the UK Department of Health, the Arthritis Research Campaign, ZonMW (the Netherlands organization for health research and development), the Health Research Board of Ireland, the UK Economic and Social Research Council, the Health and Health Services Research Fund (Hong Kong), Australian Primary Care, the Alberta Heritage Foundation for Medical Research and on a cyclical basis (in a modified form) by the Canadian Institute of Health Research

4. Technical aspects

Objectives

Buxton and Hanney identify three main reasons for undertaking an evaluation (with the Payback framework): to justify spending resources on health research; to assist with the prioritization of future expenditure; and to indicate ways to improve the conduct and management of research so as to increase the likelihood or magnitude of subsequent beneficial consequences.

Attribution

The problem of attribution of Payback to research grants or funders is acknowledged and has been explored in some depth at a 1999 international workshop on research evaluation. In addition, Buxton and Hanney acknowledge the trade-off between quality of records, the ability of researchers to recall their activities and allowing enough time for research outputs to develop.

It is also acknowledged[95] that the Payback model oversimplifies the way in which research is conducted – in particular, by abstracting from several feedback loops and secondary effects (in its logic model). At the same time it can be argued that the advantage of the Payback model is that it provides a workable framework within which to evaluate the outputs and outcomes of research.

Data collection

The Payback framework is implemented through case studies. They are based on multiple sources of evidence, whereby a number of partial sources that point towards the same conclusion are used to

---

[95] See for example Wooding, S. et al (2005): *Payback arising from Research Funding: evaluation of the Arthritis Research Campaign*, RAND Technical Report p.62

increase confidence. The main sources are: documents and literature, semi-structured key informant interviews, and bibliometric databases.

Stakeholder involvement
Evaluatees act as information sources in the Payback model. They do not have (direct) influence on the evaluation outcome. Anecdotal evidence suggests that evaluatees (by and large) agree with the evaluation outcomes.

DIUS:

1. **Introduction**

The "Economic Impacts of Investment in Research & Innovation" framework of the UK Department for Innovation, Universities and Skills (DIUS) aims to "assess the overall health of the science and innovation system, and how it delivers economic benefits".[96] It is the latest stage in a process of developing performance appraisal methods for the UK science and innovation system.

The framework is used to monitor the delivery of economic impacts at the aggregate economy level through three stages (innovation outcomes and outputs, knowledge generation, and investment in the research base) and three influence factors (framework conditions, knowledge exchange efficiency, and demand for innovation).

2. **Basic Description**

The DIUS framework is used to model the delivery of economic impacts at the aggregate economy level, through three stages (and influence factors, to be discussed later):

- Innovation outcomes and outputs (including new or improved products, processes, services; new businesses; generation of intellectual property; and wider innovation);
- Knowledge generation (in terms of adding to the stock of publicly available knowledge; and human capital); and
- Investment in the research base and innovation (including expenditure on R&D; and other forms of innovation expenditure, as defined by the CIS).

The rationale underlying the model (depicted below – Figure 19) is that the "overall economic impacts" of research are delivered through "innovation outputs and outcomes" of firms and government, who acquire and apply new ideas to provide new and improved goods and services, and public services. Innovation outputs in turn reflect the amount and quality of "investment in the research base and innovation", and "knowledge generated" by the research base.

---

[96] DIUS (2007): "Economic Impacts of Investment in Research & Innovation July 2007"; downloadable from: http://www.berr.gov.uk/files/file40398.doc
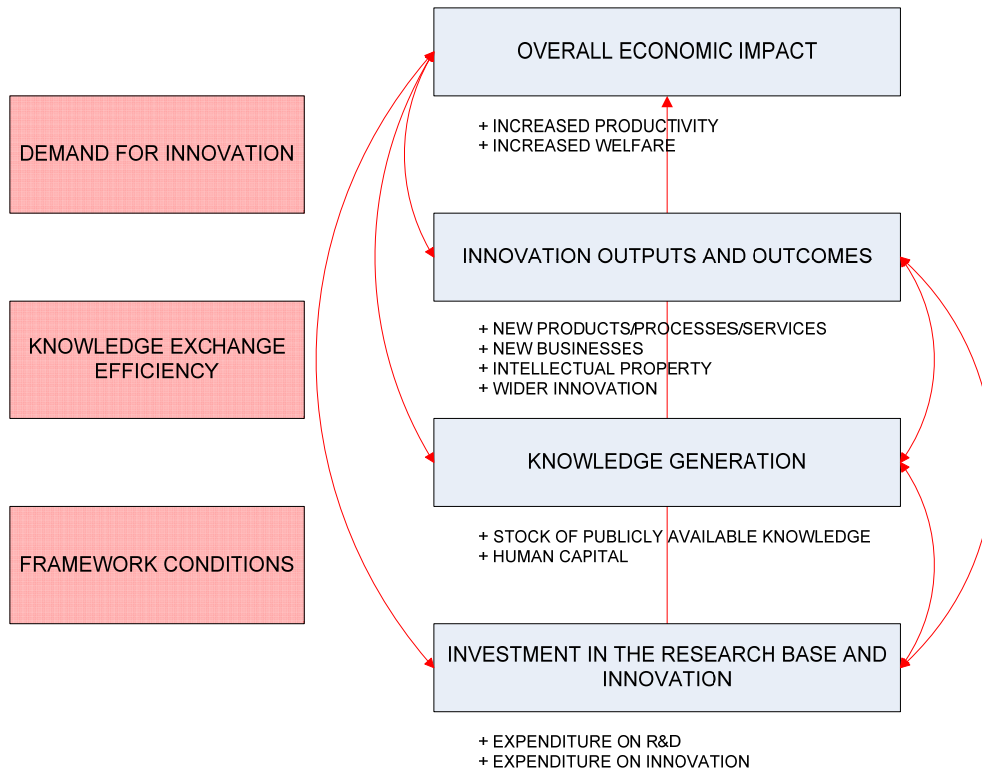
Figure 19 DIUS – Overview

"Overall economic impact" is understood in terms of "increased productivity and improved welfare". It is important to note that no (strictly) linear relationship between the individual stages is suggested (hence the arrows in the figure above). How successful the key stages (innovation, knowledge generation and investment) are in (jointly) producing economic impact depends, it is assumed, on how effectively these components work together. The main "influence factors" are:

- Framework conditions (including attractiveness of the UK to overseas investment; the intellectual property framework; public engagement; financial sustainability; and standards);
- Knowledge exchange efficiency (in terms of ease of collaboration and cooperation as well as the transit of information flows); and
- Demand for innovation (as shown in the figure above).

The different stages and "influence factors" are assessed and discussed on the basis of performance indicators and "evidence" (with the latter referring to "less frequent studies and academic research" as well as case studies). Examples (of the respective indicators and evidence) are listed in the two tables below (Tables 12 and 13).

1. Overall Economic Impact

      i. Increased productivity
- growth accounting approach to break GDP down into its sources
- Relating changes in GDP to changes in labour input, and labour productivity.

      ii. Increased welfare
- GDP figures (as broad indicators) and
- Health, environmental, social and national security outcomes (each exemplified by case study examples).

2. Innovation Outcomes and Outputs

      i. New or improved products, processes, services;
- Based on data from innovation surveys (e.g the Community Innovation Survey (CIS))

      ii. New businesses;
- Number of university spin-outs.

      iii. Generation of intellectual property;
- Patents, trademarks, registered community designs etc.

      iv. Wider innovation.
- Proportion of firms introducing organizational and/or marketing innovation (as reported in innovation surveys)

3. Investment in the Research Base and Innovation

      i. Expenditure on R&D;
- With details of proportions of publicly funded R&D, privately funded R&D, and overseas funded R&D

      ii. Other forms of innovation expenditure;
- Including expenditure on acquiring external knowledge, equipment and machinery (as defined in the CIS)

4. Knowledge Generation

      i. Adding to the stock of publicly available knowledge.
- Publication numbers and citation analysis

      ii. Human capital
- Looking at performance of UK higher education institutions, schools and further education as assessed in (independent) studies)

Table 12 DIUS – Stages

A. Framework Conditions

   i. The attractiveness of the UK to overseas investment;
   – Looking at the percentage of R&D financed by abroad and the technology balance of payments.
   ii. The intellectual property framework;
   – Performance on IP indicators
   iii. Public engagement;
   – Based on (the MORI) survey on public perception of science, science reviews (looking at science in government), media trends (in terms of coverage)
   iv. Financial sustainability;
   – Assessed mainly on the basis of biennial reviews by the funding councils.
   v. Standards
   – Based on independent studies (e.g. by the DTI)

B. Knowledge Exchange Efficiency

   i. The ease of collaboration and cooperation;
   – Based on the CIS
   ii The transit of information flows;
   – Looking at, for example, the number of patent applications filed by HEIs and the number of licences/licensing income from business, number of business representatives on governing bodies etc.

C. Demand for Innovation

   i. Demand side measures
   – Based on innovation surveys (asking, for example, about the importance of uncertain demand as a factor constraining innovation)
   ii. Business capacity
   – Again based on innovation surveys (asking, for example,, to what extent there is a lack of information on technology or lack of qualified personnel as a factor constraining innovation).

Table 13 DIUS – Influence Factors

3. Background

The framework was developed (mainly) by the UK Office of Science and Innovation in the Department of Trade and Industry, (now reorganized to form part of the Department of Innovation, Universities and Skills (DIUS)) in consultation with the former Department for Education and Skills and the UK Research Councils. In addition, input was received from key academics working in the field of evaluating outcomes of innovation and research, including SPRU and Manchester Business School. PWC and Evidence Ltd acted as consultants.

There have not been many changes since the framework was introduced in 2007 (as Annex to the annual report to the 10-year Science and Investment framework). However, the framework is the latest stage in a process of developing performance appraisal methods for the UK science and innovation system.

4. Technical aspects

Objectives

DIUS uses the framework and associated data as a way of satisfying government that its objectives are being met, and to reassure stakeholders about the health of the science and innovation system.

The indicator and other evidence that MRC and the other UK Research Councils provide are a small subset of the data and narratives prepared annually for the "Outputs Framework". The Outputs Framework is part of the "Performance Management System" that DIUS uses to oversee the work of the Research Councils.

Attribution

Problems of attribution and time lags are acknowledged: "it is highly difficult to attribute overall economic impacts [...] to the effects of a particular policy or investment". The approach deals with this problem by means of (statistical) evidence (rather than mere monitoring data) whenever possible. This, it is hoped, allows (robust) links to be established between the individual stages and between the stages and influence factors.[97]

Data collection

The framework draws on a broad set of indicators and evidence (as described above). One source of input is the UK Research Councils – which submit data and evidence for some of the categories set out in the framework. However, a considerable part of the input comes from other sources, such as government statistics and national surveys, or other studies commissioned by government.

The Research Council input to DIUS's annual report is drawn from a small subset of the data and evidence which each UK Research Council produces in an annual Outputs Framework Report. For 2006/07 the Outputs Framework reports covered all areas of the framework except for "innovation outcomes and outputs" (which relies mainly on data from innovation surveys) and the "influence factor", "demand for innovation" (which also relies mainly on data from the innovation surveys). In the case of MRC, there were some 50 quantitative or narrative indicators in the Council's 2006/07 Outputs Framework Report.

Costs

Much of the data and evidence that the MRC requires for the Outputs Framework is drawn from material the Council already gathers for other purposes. The marginal cost of preparing, collating and editing this material probably comes to less then £1k.

The preparation of data and evidence for the Economic Impacts Reporting Framework is the responsibility of DIUS.

Consequences of the evaluation

The framework informs government and other stakeholders about the health of the science and innovation system, and the extent to which government objectives are being met.

---

[97] DIUS (2007): "Economic Impacts of Investment in Research & Innovation July 2007"; downloadable from: http://www.berr.gov.uk/files/file40398.doc

**EU:**

1. Introduction

Framework Programme 7 of the European Union is meant to be a key instrument contributing to the Lisbon, Gothenburg and Barcelona objectives – the system for evaluating the programme being a vector for tracking the results of research programmes and how they are contributing to the policy goals, and a way to identify what needs to be improved so that they can be more effective in achieving these goals.

The responsibility for the evaluation of the Framework Programme rests with the evaluation unit in DG Research. It is supported by evaluation units in other DGs (JRC, INFSO, MARE, TREN, ENTR).

2. Basic Description

The Framework Programme evaluation system has been progressively updated throughout its life, but there have been moments of more radical change. One such moment was the start of Framework Programme 7. Before that, the evaluation system consisted of two main activities: annual monitoring and five-year assessments of framework programme activities. See figure below (Figure 20).

Monitoring and five-year assessments took place at two levels: at the level of specific programmes and at the level of the Framework Programme. Monitoring typically took the form of annual reviews of the progress of implementation. The reviews were conducted by expert panels. Five-year assessments were typically carried out somewhat midway through programme implementation. The idea was to combine the ex post assessment of the previous programme, the midterm appraisal of the ongoing one, and the recommendations for future activities. The five-year assessments were also conducted by expert panels.
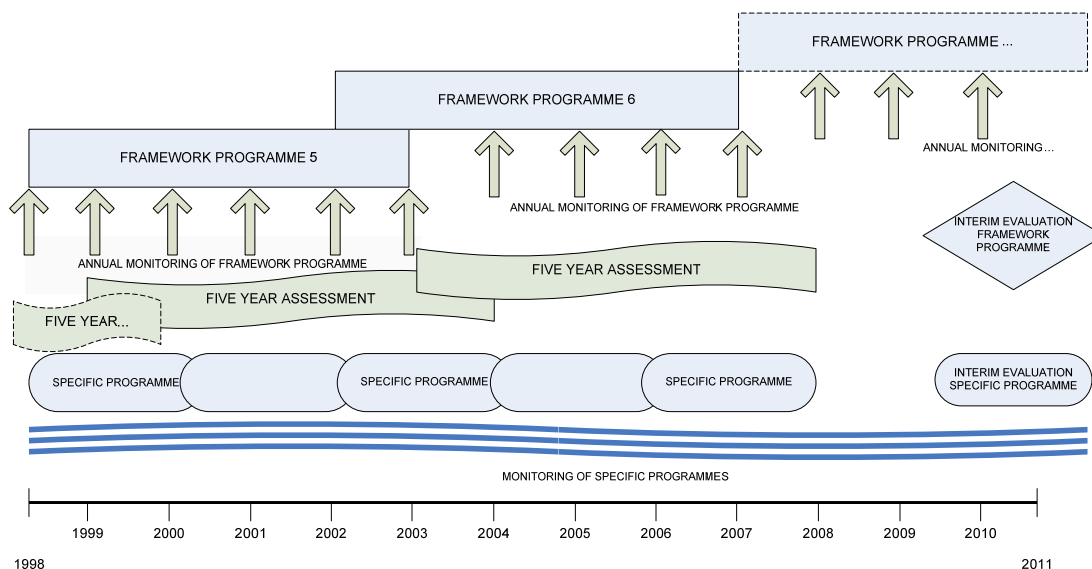
Figure 20 EU – Overview

Some elements of the old evaluation and monitoring system are still in place. These include the division into monitoring and evaluation, and framework programme and specific programme domains. The system has also continued to rely on the use of panels of high-level independent external experts (with the exception being for the monitoring exercises, which now are implemented by senior management within the Commission). What is new is:

- *The focus on "outcomes" and "impacts" and the use of "clear and verifiable objectives".* The idea is to use "a robust and coherent set of indicators to monitor achievement" with regard to (outcome and impact) objectives.

- *The concepts of an interim evaluation and ex post evaluation* (rather than five-year assessments). The interim evaluation and ex post evaluation assess the quality of the research activities and progress towards the objectives and the scientific and technical results achieved. The interim evaluation takes place 3–4 years after the start of a programme. The ex post evaluation is undertaken two years after programme completion. A table with an outline structure for possible objectives and indicators is given below.

- *The emphasis on coordinated studies*. The idea here is to develop a programme of horizontal studies for assessments of such topics as the impact of research on productivity, competitiveness and employment etc.

Programme evaluation methods include sampled analyses, case studies and longitudinal surveys, and where appropriate cost-benefit analysis and/or follow-on macroeconomic impact analysis.

	Management objectives and indicators (EC Services level)				Outcome objectives & indicators (participant level)	Impact objectives & indicators (EU level)
Framework Programme level	Budget execution rate	Time to contract	Time to payment	SME participation rate	The number of FP generated - scientific publications, citations, and their citation impact score, - new standards, tools and techniques - patent applications and licence agreements; - new products, processes and services - number of people trained through the FP - amount of energy savings and pollution reduction achieved as a result of FP research; etc.	Assessment at the aggregate FP level: - Impact on the achievement of the Lisbon, Gothenburg, Barcelona and other objectives. Assessment at the SP or project/participant level: - Contribution made to the EU S&T and economic performance (additional turnover, profit, cost savings, number of existing jobs safeguarded or new jobs created etc.)
Specific Programme 1: People (Marie Curie)	X	X	X	X	The total number (at the (sub) programme level) of - PhD participations - EU and non-EU researchers attracted (back) to the EU - Researchers that have moved from the university to the business enterprise sector; etc. The average (per project funded) number of - scientific publications and other scientific and innovative outcome	The total number of researchers exchanged within Europe, or attracted (back) from outside Europe as a result of the FP As a result of the FP the human capital gap should be reduced by X%; As a result of the FP the number of European researchers per 1000 population should reach X
Specific Programme 2: Ideas (ERC)	X	X	X	X	The average (per project funded) number of - scientific publications in SCI journals; highly cited publications; - participations by young researchers; - new tools and techniques; etc.	The total number of EU publications (plus their citations and citation impact scores) for publications that can be traced back to the FP
Specific Programme 3: Cooperation	Weighted average	Weighted average	Weighted average	Weighted average	As above	As above
Specific Programme 4: Capacities	Weighted average	Weighted average	Weighted average	Weighted average	Some of the outcomes presented above plus the number of regulations and/or directives affected by the results	A positive impact on the economic, S&T, environmental and/or social performances of the EU

Table 14 EU – Exemplary Measures

3. Background

The Commission first made public its approach to evaluation in the 1980s.[98] This was updated in 1996, when the Commission informed the European Parliament and the Council of what it then regarded as

---

[98] European Commission, Communication from the Commission to the Council on a Community plan of action relating to the evaluation of Community research and development programmes; 19.1.;1983

the relevant underlying principles for monitoring and evaluation, and set out its intended approach following the adoption of FP4.[99] From 1996 to 2006, the Commission did not fundamentally re-examine its approach to evaluation.[100]

The structure of Framework Programme evaluation activities changed significantly at the start of Framework Programme 7 (2007–2013). As described above, the new system involves a number of new exercises: An interim and ex post evaluation of each Framework Programme will replace the five-year assessment. (The evaluation of Framework Programme 6 is to be completed in 2008.) The previous panel style of annual monitoring exercise is replaced with an annual monitoring report on the implementation of the Framework Programme (by senior management within the Commission). One of the drivers for this change is the ambitious size and scope of Framework Programme 7, with its bigger budget and new instruments (ERC, technology initiatives).

One of the drivers for future change will be the 2007 report of the European Court of Auditors (CoA) on the EU research evaluation system. The CoA identified some weaknesses, such as the need for a clearer set of overall Framework Programme objectives against which evaluation could take place; for better coordination; for a more strategic planning of the evaluation activity; and for more external advice in the design of evaluations. DG Research, in collaboration with the research evaluation units in the other DGs, is looking at ways to respond to the recommendations from the CoA, in particular concerning improvements to coordination planning and the use of external advice.

4. Technical aspects

Objectives
In the Commission proposal, the objective is phrased as follows: "The programme evaluation and monitoring system supports policy formulation, accountability and learning and is essential to help improve the effectiveness and efficiency of research programmes' design and implementation."[101]

Attribution
The issue of attribution is mainly addressed qualitatively. In the survey it is asked, for example, whether participants think that their current success could be attributed "to a moderate or high extent" to the benefits accruing from their Framework Programme.

Data collection
Programme managers collect data on a day-to-day basis. But attempts are made to keep demands on participants to the (necessary) minimum. In addition, it is envisaged that a "programme evaluation data clearing house" be set up to provide a resource of information on all Community and Member States' research programme evaluations.

Costs
In the Commission Proposal it is stated: "[Evaluation and monitoring] will be resourced at a level commensurate with the challenge and comparable with international norms, taking into account the increase in size of the Framework Programme – moving towards the target of 0.5% of overall Framework budget."[102]

---

[99] Communication from the Commission to the Council and the European Parliament, "Independent external monitoring and evaluation of Community activities in the area of research and technological development"; 22.5.1996

[100] Court of Auditors (2007): Special Report No 9/2007 p.26/10

[101] Decision (2006) Concerning the Seventh Framework Programme of the European Community for Research, Technological Development and Demonstration Activities (2007–2013) p.70

[102] Ibid

Consequences of the evaluation

The information from evaluations is used in multiple ways – mostly to inform the development of new programmes and to steer existing activities. Ultimately, a poor evaluation of the Framework Programme as a whole could have serious implications on future funding levels.[103]

---

103 The CoA found however: "[…] no evidence was found that [the] findings and recommendations were taken into account for amendments to work programmes. Similarly, the DGs' ABB budgetary statements and their Annual Activity Reports do not indicate the extent to which evaluation findings were acted upon."

CDMRP*:*

1. Introduction

The Congressionally Directed Medical Research Programs (CDMRP) are part of the US Army Medical Research and Material Command (USAMRMC). The CDMRP manages (some of the) biomedical research which US Congress assigns to the USAMRMC. It was created in 1993 when Congress, in response to grassroots lobbying efforts by the breast cancer consumer advocacy community, tasked the Army with developing and managing an innovative breast cancer research programme.

The CDMRP evaluation system consists of several elements. The three main ones are: its grants management system, its product database, and its Concept Award Survey (for breast cancer research). A central element of CDMRP evaluation is that of "research product" (defined as "tangible research outcomes"). One rationale is that pressure on the CDMRP (as a military command) is even higher to develop products (rather than "just" intangibles).

2. Basic Description

Awards at the CDMRP are made in the form of grants, contracts or cooperative agreements, and the research is executed over 1 to 5 years, depending on the type of award mechanism. Each CDMRP award is assigned to a grants manager for the life of that grant, ensuring a broad knowledge of each grant, continuity among all parties involved in the award, and the most comprehensive assistance possible to the principal investigator. The grant manager (among other things) serves as the primary technical representative for the management of the award and monitors the technical progress of the overall grant.

The product database is an electronic coding system for capturing (tangible) products of funded research. The system is currently being used to catalogue and track research advances attributed to CDMRP investigators. Each product is classified according to its type, stage(s) of development, and family (group of related but different products). For an overview of the categories, see table below (Table 15). The idea of tracking research is to get a better understanding of the "impact" a certain piece of research had, but also to identify (for example) why some (initially) promising research had no impact/follow-up.

The idea of the Breast Cancer Research Program (BCRP) Concept Award is very similar. The programme is meant to support the exploration of highly innovative new concepts. The survey was designed to assess the extent to which this has any impact – for example, by providing the foundation for subsequent research.

Product Type
Animal Model – non-human animal system, such as a knockout mouse model, that mimics specific biological processes
Biological Molecule – human molecular substance, such as a gene, hormone, or protein
Biological Resource – biological material such as a cell line, used for research purposes
Clinical or Public Health Assessment – potential or tested biological procedure, such as biomarker assays and risk assessments
Clinical or Public Health Intervention – potential or tested medical and/or behavioural procedure, such as a surgical technique or diet modification programme
etc.
Stage of Development
Discovery and/or Development – initial product design, identification, and/or synthesis, or product development and/or testing in vitro systems including cell lines to determine product characteristics
Animal Validation – assessing product characteristics and effects in non-human animal models
Human Validation – preclinical assessment of product characteristics and effects in human subjects
etc.
Family
Animal Models
Biomarkers
Detection and Diagnostic Tools
Military Health and Readiness
Pharmacologic and Therapeutic Interventions

Table 15 CDMRP – Product Database

3. Background

The evaluation efforts (outlined) in the CDMRP are coordinated by an evaluation division. It was established in response to an assessment of the Breast Cancer Research Program (BCRP) by the IOM. The IOM was asked to include a review of the portfolio of funded research, assess programme management and achievements, and recommend areas for funding that have not been funded or areas that need additional emphasis.

As noted in the CDMRP 2005 annual report, "[t]he result of this review was a report published in 1997 that concluded with 3 major and 13 secondary recommendations. One of the major recommendations was that the CDMRP "develop and implement a plan with benchmarks and appropriate tools to measure achievements and progress towards goals of the BCRP both annually and over time." In addition, "the CDMRP is accountable for the expenditure of congressional appropriations – accountable for the consumer advocacy groups, to the scientific community, to Congress, and to the American public at large".[104]

Currently the evaluation division of the CDMRP is developing and refining analysis techniques for its database.

---

[104] The report can be found under www.cdmrp.army.mil/annreports/2005annrep/default.htm

4. Technical aspects

Objectives

"The continuation of the CDMRP is dependent upon annual congressional appropriations. The CDMRP, in turn, has an obligation to demonstrate adherence to congressional mandates, verify return on investment, and keep stakeholders – Congress, the DOD, and the public – apprised of achievements and ongoing activities."[105]

Costs

As for the database, grant-holders are required to write progress reports, which are then mined by the evaluation division for products. This is estimated to take around 15 minutes per report. Follow-up typically takes another couple of hours.

Communication of results

The results of the various evaluations are disseminated by means of an annual report (in the form of research highlights) through the CDMRP website. In addition, "consumers" (who are typically survivors and their families) are invited to attend multidisciplinary meetings held by the CDMRP (such as the Breast Cancer Research Program's Era of Hope meeting) where they can learn about the scientific advances (through CDMRP funding)

---

[105] CDMRP (2005):Annual Reort- The report can be found under www.cdmrp.army.mil/annreports/2005annrep/default.htm

Appendix B: The Canadian Landscape for health research

Despite a common goal to improve the health of Canadians, the Canadian health research system is a diverse entity influenced by many factors. It comprises a variety of stakeholders, each with different goals and strategies to obtain these goals. These stakeholders must be examined separately in order to understand how the system functions.

Health research is a global enterprise within which individual countries compete and benchmark against each other – yet it is also one in which countries collaborate and share research findings. For this reason, it is important to consider the Canadian system within the international context.

The Canadian Health Research Landscape

An overview of the funding and performing sectors of the health research field is outlined below. This is followed by summaries of major Canadian granting agencies and their funding programs and outputs.

In 2007, Canada spent an estimated $160.1B on healthcare (Canadian Institute for Health Information 2007). In contrast, expenditure on the health research system was $6.3B (Science, Innovation and Electronic Information Division 2008). In theory, this research funding should help to reduce the healthcare funding. To understand how health research funds are spent in Canada it is important to analyze the system of funders and producers of research in the country.

Perhaps the best way to examine the Canadian health research system is to first break it into its various sectors: the Public Sector, the Private Sector, and the International Sector. Within each sectors there are different groups or types of funders:

1. **Public Sector**
   a. Federal Funders
   b. Provincial Funders
   c. Higher Education
2. **Private Sector**
   a. Industry/Business Enterprise
   b. Private-non-profit
3. **International Sector**
   a. Foreign Investment

These groups are not isolated from one another and their relationships can be complex. For example, the funding for Higher Education institutions comes mostly from provincial government coffers, while industry and the Federal Government have joint funding streams. In some cases Federal Funders provide lump sum funding endowments to non-profit organizations such as the Canadian Health Services Research Foundation (CHSRF).

There are different types of activities supported by research funders. Research activity and human resources receive the majority of funds (allocated through organizations such as CIHR) but infrastructure is also supported (through organizations such as the Canadian Foundation for Innovation (CFI)).

Figure 1 shows the six groups of the Canadian health research system and their interaction with the different stakeholders responsible for facilitating funds to researchers and organizations. For reasons of maintaining simplicity, only major funders are shown.
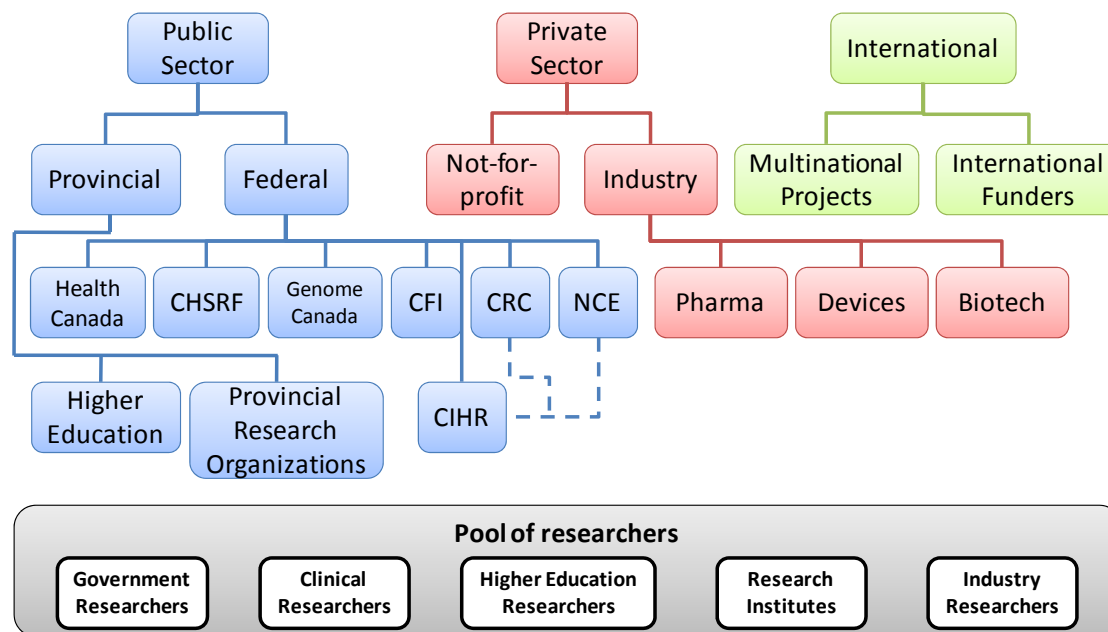

Figure 21. Major funders and funding flows in the Canadian health research system (Nason 2008)

In the Public Sector, there are several federal bodies that are directly funded by the Federal Government (Health Canada, CIHR); bodies funded through federal endowment funds (CHSRF), for infrastructure (CFI), and research programs that are not exclusively restricted to the area of health research (NCE, CRC). It should be noted that the Networks of Centres of Excellence (NCE) and the Canadian Research Chairs (CRC) are jointly funded through the three federal research councils, with CIHR providing funding in the health realms of each research initiative – as shown by the dotted lines connecting CIHR to the CRC and NCE funding (Nason 2008).

Of note, 'Foreign Investment' is a group listed under the International Sector. In Figure 1, 'Foreign Investment' is broken into Multinational Projects and International Funders. An example of a Multinational Project would be the Human Genome Project, which relied on collaboration between countries. An example of an International Funder would be the National Institutes of Health (NIH) in the United States, which funds a significant number of research projects in Canada.

At the bottom of Figure 1 (Pool of researchers) are the groups who perform research in Canada. Researchers can be funded by any of the above listed funding groups, although Government Researchers will not receive extramural funding.

In 2007, the Higher Education Sector was the major funder of research in Canada, followed by the Business and Federal Government sectors. The major performers of research are Higher Education, with most of the remaining research performed by Business (Figure 2).
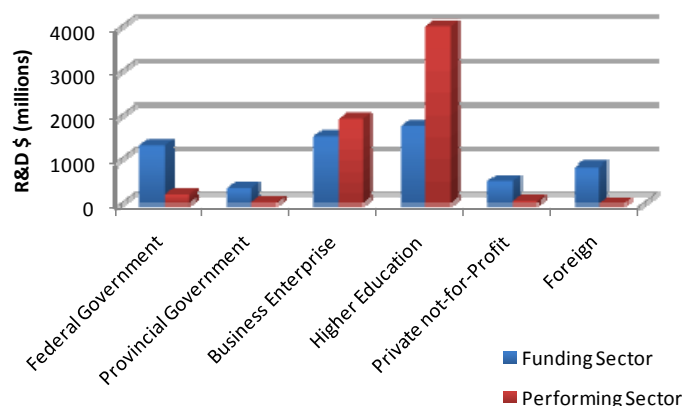
Figure 22. Funding and performance of health research in Canada in 2007 (Science, Innovation and Electronic Information Division 2008)

Within each sector shown in Figure 21 (Public, Private, and International), there are funding agencies who want and need to understand what impacts their research funding has had (or will have) and how impacts arise. The diversity of sponsors of this Assessment demonstrates the common desire of all research funders, regardless of sector, to understand the impacts arising from the research they fund.

We will now go into more detail in describing each of the six groups of the Canadian health research system referred to in Figure 22 (Federal Funders, Provincial Funders, Industry, Higher Education, Private-not-for-Profit, and Foreign Investment), the categorization used by Statistics Canada in data gathering on funding. In Appendix C we highlight funders from each sector to demonstrate how they evaluate their performance.

Public Sector: Federal Funders

In the last decade, several federal research initiatives have been created, including the Canadian Institutes of Health Research (CIHR), Canada Research Chairs (CRC), the Networks of Centres of Excellence (NCE), Genome Canada (GC), the Canadian Foundation for Innovation (CFI), Genome Canada, and the Canadian Health Services Research Foundation (CHSRF). These initiatives are funded by the Federal Government.

The Federal Government runs both intramural and extramural research programs. Intramural research (internal research conducted in government designated laboratories) accounted for 18 percent ($195M) of Federal Government health research expenditures in 2004. The majority of health research funded is extramural research (carried out by non-federal organizations), which received the remaining 82 percent ($906M) of 2004 spending (Madore, O., Norris, S. 2006). In 2007, the Federal government spent $1.3B on R&D in the health field (Science, Innovation and Electronic Information Division 2008).

Federal Intramural Research Programs

Federal expenditures in Intramural Research and Development have remained fairly stable from 2000/2001 to 2006/2007, rising from $4B to $4.9B (Statistics Canada 2007).

The Federal Government's primary organization for R&D is the National Research Council of Canada (NRC). The NRC is made up of over twenty institutes and national programs that are located throughout Canada. These institutes and programs are divided by area: Life Sciences; Physical Sciences; Engineering; Technology and Industry Support; and Corporate Services.

The NRC is a large program: for example, the Biotechnology Research Institute (BRI) of the NRC is the largest Canadian facility devoted to biotechnology (National Research Council of Canada 2008). The National Research Council (NRC) was funded by about 26 percent of Federal Government expenditures in Intramural R&D in 2007 (Statistics Canada 2007).

Federal agencies are considered intramural entities, and include the Canadian Institutes of Health Research (CIHR), the Canada Foundation for Innovation (CFI), Genome Canada (GC), Health Canada, the Natural Sciences & Engineering Research Council (NSERC), the Public Health Agency of Canada (PHAC), and the Social Sciences & Humanities Research Council (SSHRC). We will now describe in greater detail some of the most prominent federal funding entities, of which the Canadian Institutes of Health Research is the largest.

Canadian Institutes of Health Research (CIHR)

Budget: $962.5M (2008-2009)(Canadian Institutes of Health Research 2008)

There are three federal granting agencies in Canada:

> **CIHR's mandate is: 'To excel, according to internationally accepted standards of scientific excellence, in the creation of new knowledge and its translation into improved health for Canadians, more effective health services and products and a strengthened Canadian health care system' (Canadian Institutes of Health Research 2007)**

the Canadian Institutes of Health Research (CIHR), the Natural Sciences and Engineering Research Council (NSERC), and the Social Sciences and Humanities Research Council (SSHRC). Each of these granting councils is annually allocated funds by Parliament. NSERC and SSHRC are accountable to Parliament through the Minister of Industry, while CIHR is accountable through the Minister of Health.

CIHR is relatively new compared with the other two granting councils (both formed in the 1970s) and was created as the successor to the Medical Research Council of Canada (MRCC) in 2000 under the CIHR Act. The MRCC was established in 1960 and its recognition increased later in the decade, when more money began to be spent on health-care issues. In 1992, in response to a desire for an increased research scope, the board of the MRCC decided to become a part of the (not yet existent) CIHR (Nason 2008). Today CIHR has grown into the largest federal funder of health research in the country. It is comprised of thirteen 'virtual' institutes (not necessarily funding co-located researchers) funding specific areas of health such as Aboriginal Peoples' Health, Genetics, and Population and Public Health.

As mentioned above, CIHR is the largest federal funder of health research; its expenditures accounted for nearly 60 percent of total federal investment in health research in 2004-2005 (Madore, O., Norris, S. 2006). Both NSERC and SSHRC fund components of health research, but only as it relates to their primary mandates and/or through joint programs such as the Collaborative Health Research Program (NSERC and CIHR) or other tri-agency initiatives. CIHR allocates funds to the Canada Research Chairs (CRC) program as well as the health related Networks of Centres of Excellence (NCE). The majority of funding (70 percent) is allocated to investigator initiated research, with the remainder going to support targeted research funded by the Institutes and/or corporate CIHR programs (Canadian Institutes of Health Research 2007) .

Funding is organized by dividing health research into four categories or 'pillars': biomedical; clinical; health systems and services; and population and public health (see Box 1). This classification of research areas has become common nomenclature in health research organizations across the country.

Box 1. Four pillars of health research (Canadian Institutes of Health Research 2007)

Pillar I	Biomedical Research
Pillar II	Clinical Research
Pillar III	Health Systems and Services Research
Pillar IV	Population and Public Health Research

Pillar I research is the pillar in which most infrastructure and HQP are required (lab equipment and personnel, etc.). When applying for funds, researchers must identify the pillar they associate themselves and their research with; sometimes there is a possibility that research pillars might overlap. Of note, Pillars III and IV have been receiving larger increases in funding year on year since CIHRs inception in 2000 (Canadian Institutes of Health Research 2007).

Framework/Indicators

Both domestically and internationally, CIHR has been an instigator in the quest for a suitable framework that could be applied within Canada to evaluate the impact of health research. In 2005, 'A framework for measuring the impact of health research' was published (Canadian Institutes for Health Research 2005). This work built upon the Buxton-Hanney Payback Model (Buxton, Hanney 1996). In 2006, CIHR revised this model (Bernstein A, Hicks V, Borbey P, Campbell T 2006) and in 2008, it updated it again (CIHR 2008). Currently CIHR breaks up the impact of health research into five categories:

1. Advancing Knowledge
2. Capacity Building
3. Informing Decision Making
4. Health Impacts
5. Economic Impacts

This framework and accompanying logic model is examined in more detail in Appendix C.

Canada Research Chairs (CRC)

Budget: ~$300 M/year (Canada Research Chairs 2008)

In 2000, the Government of Canada established the Canada Research Chairs (CRC) program with the goal of creating 2,000 research professorships in Canadian universities by 2008. The initiation of this program was in response to the changing knowledge-based economy within Canada and other G-7 nations, and the recognition that Canada needed new programs to attract top researchers to domestic universities. The federal government provided $900 M to support the start-up of the program and to fund the first 2,000 Chair holders. The CRC program is close to achieving its goal: as of June 2008, there were 1,829 active Canada Research Chairs.

The CRC program strives to increase Canadian research capacity, improve the training of Highly Qualified Personnel (HQP), improve universities' capacity to generate and apply new knowledge and use research resources in the best ways possible (Canada Research Chairs 2006).

Chairs are awarded in the natural, health and social sciences through the three federal granting agencies (CIHR, NSERC, and SSHRC). Chair holders are also eligible to receive infrastructure report from CFI, which made an initial partnership with the CRC to provide $250 M in infrastructure support for the first 2,000 Chair holders. CFI created the New Leaders Opportunity Fund (LOF) in 2005 as a follow-on program to its initial pledged support (this program still funds infrastructure).

Chairs are divided into two tiers:

Tier 1 Chairs: are tenable for seven years and renewable. They are awarded to outstanding researchers acknowledged by their peers as world leaders in their field. For each Tier 1 Chair a university has, it will receive $200,000 annually for seven years. As of June 2008, 801 Tier 1 chairs had been awarded.

Tier 2 Chairs: are tenable for five years and renewable once. They are awarded to exceptional emerging researchers acknowledged by national experts as having the potential to lead in their field. For each Tier 2 Chair a university has, it will receive $100,000 annually for five years. As of June 2008, 1,028 Tier 2 chairs had been awarded.

It is important to distinguish the Canada Research Chairs program from the more recently established Canada Excellence Research Chairs Program (CERC). The latter program was created as a specific response to the Government of Canada's Science and Technology Strategy (Fast 2007), in that it funds researchers in the four areas highlighted in the Strategy: environmental sciences and technologies; natural resources and energy; health and related life sciences; and information and communication technologies. The CERC program is intended to go one step further than that of the CRC; it specifically targets the priorities outlined by the Federal Government as being of economic importance to Canada (Canada Excellence Research Chairs). Like the CRC, this is a tri-agency funding effort, but its secretariat is housed in SSHRC.

Framework/Indicators

In March 2006, the CRC developed an 'Integrated Results Based Management and Accountability Framework and Risk - Based Audit Framework' to describe performance measurement and evaluation strategies, as well as to discuss the risks that could affect the performance of the program. This document provides a background to the CRC and lists the various beneficiaries of the initiative (e.g., Canadian universities and their research affiliates, Chair holders, Trainees).

The Framework includes a logic model; outlines the roles, responsibilities and relationships of the program; how the program is to be monitored; and the evaluation strategy used by the CRC (Canada Research Chairs 2006).

Networks of Centres of Excellence (NCE)

Budget: $82.4 M (2006-2007) (Networks of Centres of Excellence)

The Networks of Centres of Excellence (NCE) is a permanent program that funds Centres that work to develop partnerships between universities, industry, governments, and not-for-profit organizations. The NCE is jointly funded by NSERC, SSHRC, CIHR and Industry Canada. The NCE is one of the four initiatives run by the NCE Secretariat, established in 1989 (the others are the Centres of Excellence for Commercialization and Research; Business-Led Networks of Centres of Excellence; and the Industrial Research and Development Internship Program).

The primary objective of the program is to translate Canadian research and entrepreneurial talent into economic and social benefits for all Canadians. These nation-wide multidisciplinary and multi-sectoral partnerships connect researchers with those savvy in industry and strategic investment.

Networks can be supported for two seven year cycles. Currently, out of the fifteen total networks that exist, there are seven in the area of Health, Human Development and Biotechnology. These include the Canadian Stroke Network (University of Ottawa) and the Canadian Arthritis Network (Mount Sinai Hospital, Toronto). Health-related research funded through the NCEs accounts for $27.5M of its budget (Networks of Centres of Excellence).

> In an average NCE cycle, the Networks produce approximately: 27,000 refereed publications; train 2,300 new trainees; file 320 patents; negotiate 540 licenses; and launch 70 spin-off companies (Networks of Centres of Excellence 2008)

Between 2006-2007, the NCE Program supported more than 6,000 researchers and HQP. The networks partnered with 830 Canadian companies, 333 provincial and federal government departments, and 584 agencies from Canada and around the world (Networks of Centres of Excellence 2008).

In 2006-2007, in addition to their annual budget, the Networks stimulated outside cash and in-kind investments totalling almost $59 M, including more than $22M by the participating private sector companies (Networks of Centres of Excellence 2008).

Framework/Indicators

In 2002, the NCE produced a 'Results-Based Management and Accountability Framework' (Networks of Centres of Excellence 2002a). A logic model was developed as part of this process, which led to the development of indicators and an evaluation and reporting strategy. Additionally, a 'Risk-Based Audit Framework' (Networks of Centres of Excellence 2002b) was drawn up the same year. Both can be found on the NCE website: www.nce.gc.ca.

Genome Canada (GC)

Budget: $1.5 B/year (Government of Canada: $700 M/ Partners: $800 M) (Genome Canada 2008b)

Established by the Government of Canada in 2000 (and funded through Industry Canada), Genome Canada (GC) is the primary funding and information entity related to genomics and proteomics within Canada. Genome Canada has established six Genome Centres across the country (Atlantic, Quebec, Ontario, Prairies, Alberta, and British Columbia) with the aim of becoming a world leader in genomic and proteomics research. Together with its six Centres and other partners (industry, government departments and agencies, universities, and research hospitals) Genome Canada invests in and manages large-scale research products in key areas such as agriculture, the environment, fisheries, forestry, health and the development of new technology. GC also supports research projects aimed at studying and analyzing the ethical, environmental, economic, legal and social issues related to genomics research (GE3LS) (Genome Canada 2008a).

In March 2004, Genome Canada produced an 'Interim Evaluation of Genome Canada' (Bearing Point 2004). According to this report, 57 percent of all funding was in the area of health.

Framework/Indicators

Genome Canada has recently developed a 'high level framework' called 'The Performance, Audit and Evaluation Strategy'. This framework was first implemented for fiscal year 2007-08 (Genome Canada 2007) and can be accessed via the GC website: www.genomecanada.ca.

Canadian Foundation for Innovation (CFI)

Budget: Differs per year (In 2007 the Government of Canada allocated $510 M to CFI for 2007-2010)(Honourable James M. Flaherty, Minister of Finance 2007)

Created by the Government of Canada in 1997, the Canadian Foundation for Innovation (CFI) funds research infrastructure in Canadian universities and their affiliated hospitals and research institutions. Research infrastructure includes equipment, buildings, laboratories and databases (Canada Foundation for Innovation 2005).

The primary mandate of CFI is to strengthen the capacity of these institutions by supporting them to carry out world-class research and technology development that will benefit Canadians. Since its inception, CFI has received $3.65B from the Government of Canada (Canada Foundation for Innovation 2005). CFI is committed to funding up to 40 percent of a given project's infrastructure costs; the remainder is financed by a combination of public, private and voluntary organizations. According to CFI's 2004-2005 Annual Report, by 2010, the total capital investment by the CFI and its partner research institutions is estimated to exceed $11 B (Canada Foundation for Innovation 2005).

The research enabled by CFI support is said to be creating the conditions necessary for sustainable, long-term economic growth, including the creation of spin-off ventures and the commercialization of discoveries.

CFI categorizes research into the areas of health sciences, human and social sciences, natural sciences and engineering, arts and literature, and multidisciplinary research. Out of the 84 projects listed in CFI's project database, (for all years included) under Multi-disciplinarity 12 were classified as belong to the field of 'health' (Canada Foundation for Innovation).

Framework/Indicators

CFI does not have a publicly available framework or logic model. However, their 2005/2006 Annual Report illustrates case samples of social and economic benefits. It also details its accomplishments, such as number of spin-off companies created and number of new products developed (Canada Foundation for Innovation 2006).

Indirect Costs Program (ICP)

Budget: $315M/year (Government of Canada 2008a)

The Indirect Costs Program (ICP) is a federal program that assists universities and colleges by funding the 'indirect' costs of administering and managing research activities. Examples of indirect costs include renovating laboratories or upgrading computer systems. The ICP works alongside the funding programs of CIHR, NSERC and SSHRC to ensure that as much of a research project as possible is funded. This program assists over 125 Canadian postsecondary institutions (Government of Canada 2008a) and is administered by the secretariat of the CRC Program (housed at SSHRC).

Framework/Indicators

There is no publicly available evaluation strategy used by the ICP. There is an Interagency Program Review Committee, however, which meets on an annual basis. The Secretariat meets with the vice-presidents of programs from CIHR, NSERC and SSHR to review new programs (Government of Canada 2008a).

Canadian Health Services Research Foundation (CHSRF)

Budget: $15-16 M /year (Canadian Health Services Research Foundation n.d.)

Established in 1997, the Canadian Health Services Research Foundation (CHSRF) is an independent, not-for-profit organization, supported by endowed funds from the federal government. CHSRF funds health services and systems research and provides research opportunities to both researchers and decision makers to investigate specific-health-system questions. It also provides training for senior decision

makers (through the EXTRA program) in nursing, medicine and health administration in learning how to find and apply research and to facilitate evidence-based decision making. CHSRF is a broker in knowledge translation activities supporting health systems and services research.

CHSRF works has an extensive array of partnerships, including: the Commonwealth Fund, the Netherlands Organisation for Health Research and Development (ZonMW), CIHR, and other national and provincial organizations.

Framework/Indicators

CHSRF has developed logic model, however it is an internal tool and not available publicly. The outputs of research funded by CHSRF are available on their website, www.chsrf.ca.

Public Sector: Provincial Funders

Provincial funders are the second group that fund research in the Public Sector. Contributions to health research by the provincial governments were $0.4B in 2007 (Statistics Canada 2008a). Some provinces have created agencies to fund capacity building, grants or commissioned research. These include (but are not limited to) the Alberta Heritage Foundation for Medical Research (AHFMR), the Michael Smith Foundation for Health Research (MSFHR), the Nova Scotia Health Research Foundation (NSHRF) and la Fondation de la recherché en santé du Quebec (FRSQ).

Framework/Indicators

Provincial research organizations are at various stages in terms of employing frameworks and evaluation methods to assess their research. In most cases, reports and publications provided by these provincial health research organizations provide only examples of returns on investment.

Provincial organizations are working together, however, and make up the National Alliance of Provincial Health Research Organizations (NAPHRO) (Beaudet 2007).

Higher Education

In 2003, the Higher Education Sector (Canadian universities and their affiliated hospitals and institutions) surpassed the Business/Industry Sector as the leading funder of health research in Canada. This sector is both the top funder and top performer of health research in Canada and in 2007 spent $1.8B (Statistics Canada 2008a). As mentioned above, the Federal Government is the third largest funding sector in the country, contributing one-fifth of funds in R&D in health in 2007 (Science, Innovation and Electronic Information Division 2008).

It may be surprising to some that the Higher Education Sector spends the most on health research, but the sector funds research through faculty salaries, space, infrastructure etc. One factor to consider in the volume of HE research funding is the success of some Canadian universities in significant fundraising. For example, from 1997 – 2004, the University of Toronto received $1B in donations. The campaign drive involved the entire university and 80 percent of funds raised went to human capital (to fund chairs, professorships, student aid, and academic programs); the remaining 20 percent supported capital projects. The Dean of the Faculty of Arts & Sciences wrote that 'the campaign has… trigger(ed) the largest capital expansion program in more than 40 years' (University of Toronto, n.d.). Government matching programs played a key role in the success of the campaign. The University of Toronto was supported by other initiatives including the Ontario Student Opportunity Trust Fund, the Ontario Graduate Scholarship and Ontario Graduate Scholarship in Science and Technology programs, Superbuild, the Ontario Innovation Trust, and the Canadian Foundation for Innovation (University of Toronto, n.d.).

Another example is McGill University, which launched 'Campaign McGill: History in the Making' in October 2007, the most ambitious fundraising drive in its history, scheduled to run until 2012. It is seeking to raise $750M for the university and as of March 2008, had already collected $398.2M. These programs are targeted and are focused on raising funds for key areas that have been identified as important to the higher education sector in the 21st century. For instance, McGill plans to use the money raised to help address five key challenges: advancing health and wellness; building global prosperity; furthering the next generation of science and technology; creating environmental sustainability; and strengthening culture and civil society (McGill University Spring 2008).

Framework/Indicators

There is no one overarching framework that is used by all higher education institutions in Canada. Most universities do have their own way of measuring goals and performance. The Association of Universities and Colleges of Canada (AUCC), a non-governmental, not-for-profit organization that represents 92 Canadian public and private-not-for-profit universities and university degree level colleges published 'Momentum' in 2005, which outlined the benefits Canadians would gain from investing in research.

Data on the outputs of university funded research is available through Statistics Canada.

Private Sector: Industry/Business Enterprise

In 2007, the business enterprise sector funded $1.5B of R&D in the health field into Canadian health research (Statistics Canada 2008a). This sector is comprised of, but not limited to, industries in biotechnology, pharmaceuticals and medical devices. Overviews of these three sectors are provided below.

Biotechnology

In 2005, Canada had 532 biotech companies, 303 of which were classified in the Sector of Human Health (Industry Canada 2008a). In 2005, the revenue generated by the biotechnology industry amounted to $4.2B. Between 1997 and 2005, Canadian companies (on average) have increased their R&D spending by 39 percent each year (Industry Canada 2008a).

Among the G7 countries, Canada has established the fastest rate of growth in the number of workers devoted to R&D in external patent applications and in business expenditures on R&D. Canada's growing R&D capacity in biopharmaceuticals is an evolution from a research base that has won an international reputation in fields such as genomics, proteomics, bioinformatics, stem cells, immune-therapies, and protein engineering and new drug delivery systems. Sixteen Canadian universities are affiliated with a network of more than 100 teaching hospitals and research institutes worldwide.

Canada has continued to have the second highest number of biotechnology companies in the world demonstrating a supportive business climate and Canada's commitment to growing this vital sector (Industry Canada 2008a).

Pharmaceuticals

The Canadian pharmaceutical industry is the second largest in the world in terms of industry size. It comprises 130 pharmaceutical companies (Government of Canada 2007). It also has 165 biotechnology companies.

In 2005, the pharmaceutical manufacturing industry employed about 40,000 people directly and a further 35,000 indirectly. This is an industry that continues to grow: according to the Government of Canada, Canada is the third fastest growing pharmaceutical market in the world, with a growth rate of 8.9 percent per year. Canada is the eighth largest pharmaceutical market in terms of sales (Government of Canada 2007).

In 2006, intramural research accounted for 50.5 percent of expenditures (this declined from the 52.6 percent reported in 2005). Research performed by other companies on behalf of patentees rose to 22.1 percent of the current R&D expenditure (in 2006), while the combined share of universities and hospitals was 16.2 percent.

Of note, Canadian pharmaceutical exports have grown 115.4 percent (from 1998-2003) [(Government of Canada 2007).

Framework/Indicators

Rx&D Canada's Research Based Pharmaceutical Companies provides statistics on how the pharmaceutical industry has saved lives and reduced hospital says. Further information can be found on their website, at http://www.canadapharma.org

Medical Devices

The manufacturing of medical devices involves the application of biomedical and engineering disciplines. The Canadian medical device industry benefits from the strengths of associated Canadian industries including biotechnology, advanced materials, microelectronics, telecommunications, and software and informatics. The industry is able to draw upon world-class innovative research being conducted in Canadian universities, research institutes and hospitals. Nearly 10 percent of Canadian medical device firms are spin-offs from universities, other firms or laboratories (Industry Canada 2008b). Additionally, a number of medical device firms work in collaboration with other organizations (such as universities, hospitals, smaller or larger firms, government departments, etc.).

Canada has approximately 800 manufacturers and distributors of medical devices. Like the growth in pharmaceuticals, medical devices sales to international markets have increased approximately 250% over the last decade (Government of Canada 2007).

From 2004–2005, the medical device manufacturing and development industry consisted of 1,101 facilities, comprising approximately 998 firms. (An additional 685 facilities, operated by 602 firms, solely engaged in distribution are not included in the following analysis). Medical device manufacturing and development facilities were generally smaller in size than pharmaceuticals, as more than half (57 percent) had fewer than 25 employees and 37 percent had between 25 and 49 employees. Of the remaining facilities, only 45 (4 percent) were of medium size (50-150 employees), and less than 1 percent were large (greater than 150 employees). Medical device-related employment in 2004 – 2005 rose to approximately 26,000 (compared to 22,000 in 2000 (Government of Canada n.d.). Approximately 90 percent of the medical device facilities were Canadian owned; a number unchanged from 2000. Foreign-owned facilities tended to be larger as 21 percent had 50 or more employees, compared to just 4 percent of domestically-owned facilities.

The Medical device Sector is supported by the National Research Council, NSERC, CIHR, CFI, among other federal and regional players.

Private Sector: Not-for-Profit Organizations

There are an estimated 175,000 not-for-profit organizations in Canada, 5,500 of which are primarily dedicated to health (Statistics Canada 2005). Approximately $0.5B was allocated to health R&D in 2007 by private non-profit organizations (Statistics Canada 2008a). These organizations vary in size – some are extremely large and well-known such as the Heart and Stroke Foundation and the National Cancer Institute of Canada (who together contributed $108M to health research in 2006 (Heart and Stroke Foundation of Canada 2006, National Cancer Institute of Canada 2007). However, the majority are smaller, regional organizations.

Unfortunately, due to the size and variety of this sector, it is difficult to obtain consistent data for it. This sector is at various stages of development in terms of frameworks and indicators.

International Sector: Foreign Investment

Foreign investments in health research accounted for about 13 percent of the total expenditures on R&D in the health field in 2007, amounting to approximately $833 M (Science, Innovation and Electronic Information Division 2008). Close to 93 percent of these funds were allocated to the Business enterprise sector, 6 percent to Higher Education, and 1 percent to the private-non-profit sector (Science, Innovation and Electronic Information Division 2008).

The US National Institutes of Health (NIH) is the primary foreign investor in the Higher Education sector in Canada. In 2007, 195 grants totalling $44 M were granted to Canadian researchers, which accounted for 31 percent of all NIH foreign grants (National Institutes of Health 2006). Over the past ten years, this number has been rising. It should be noted that the NIH only fund external research where it feels the work is of strategic importance to its mandate – and where the best infrastructure and expertise for a project.

Canada's Place in the World

Since health research is a worldwide enterprise, it is necessary to place the Canadian system within a global context.

We will start by examining funding inputs into the Canadian system by looking at health R&D, total R&D, healthcare expenditure and GDP. Funding input is a simple and reliable measure to make comparisons and contrasts with. We will then compare Canadian health research expenditures with expenditures of select OECD countries.

Inputs to Health R&D in Canada and Select OECD Countries

Health research in Canada receives less than 5 percent of the funding that healthcare does (Canadian Institute for Health Information 2007, Science, Innovation and Electronic Information Division 2008). However, health R&D is a significant proportion of total R&D spending in Canada, accounting for just over a fifth in 2007 (Science, Innovation and Electronic Information Division 2008). Canada's GDP in 2007 was estimated at $1.54 trillion (Statistics Canada 2008c); therefore the money spent on health R&D was 0.4 percent of GDP.

When examining research funding within a country, the following measures are used: Gross Expenditure on R&D (GERD); Business Expenditure on R&D (BERD); Government Budget Appropriations or Outlays on R&D (GBAORD); and the spending by the pharmaceutical industry on R&D (Organisation for Economic Co-Operation and Development 2007).

This data is collected by the OECD and used by governments to monitor funding inputs (Treasury Board of Canada Secretariat 2007); by research organizations to indicate Canada's R&D spending relative to other countries (The Conference Board of Canada 2007, The Conference Board of Canada 2004); and by the press to illustrate how Canada values R&D (Wahl 2008).

Canadian GERD as a percentage of GDP is just below 2 percent. This is higher than the UK and OECD average, but lower than the USA and Germany (Organisation for Economic Co-Operation and Development 2008).

Canadian BERD (2008) is estimated to be $16.3B (Statistics Canada 2008d), which is just below 1% of the GDP (Statistics Canada 2008b). This BERD level is below the OECD average, which is around 1.5% of GDP (Organisation for Economic Co-Operation and Development 2007). However, despite having a considerably low BERD, the amount of higher education funding of R&D is significantly higher than other countries and is second out of the OECD countries only to Sweden (Government of Canada 2008b).

Narrowing down the scope to examine proportions of GDP spent on health research is more difficult, since health research is widespread across a range of disciplines. As mentioned above, the OECD uses Government Budget Appropriations or Outlays on R&D (GBAORD) as a proxy to identify what government spends in health-related research.

GBAORD in Canada was $1.3B in 2007 (equivalent to less than 0.1% of GDP). In contrast, in the USA it was 0.22 percent in 2005 and 0.03 percent in Japan in 2006 (Organisation for Economic Co-Operation and Development 2007). However, GBAORD only includes government expenditure; it does not capture the contributions from the education and business /industry sectors, which are two of the largest funders of health research in Canada.

Narrowing down the focus into health expenditures, we find that Canada spent a total of $6.3B in 2007. Approximately a third of this came from the higher education sector, a third came from the business sector, and the remainder came from federal, provincial and not-for-profit funders (Science, Innovation and Electronic Information Division 2008). Of the R&D carried out in Canada in 2007, health represented over one fifth of the R&D funding for the country (Science, Innovation and Electronic Information Division 2008).

For example, when examining industry funding, the OECD uses expenditures made by the pharmaceutical industry as a proxy for funding in health R&D by business. In 2007, the Canadian business sector spent $1.5B (around 0.1% of GDP). Compare this with 0.5% of GDP in Sweden and over 0.2% in the UK (2004 figures) and around 0.15% in the USA (2003 figure) (Organisation for Economic Co-Operation and Development 2007).

Although the above figures are illuminating, the very fact that there is a need for an evaluation of research impacts tells us that these input figures do not provide anything more than information on how much a country values health research. The use of these measures relies on the assumption that if health research provides a positively high impact, then the market will invest in it. This, however, oversimplifies the stresses and pressures governments face when dealing with allocation of funds. If there were an agreed 'right amount' to spend on health research (as a percentage of GPD, per capita, per dollars spent on healthcare, etc.), then these input values might hold greater analytical power.

Of course, there is no 'right amount'; an illustration of this is the wide diversity in the percentage of GDP spent on research in OECD countries (Figure 23).
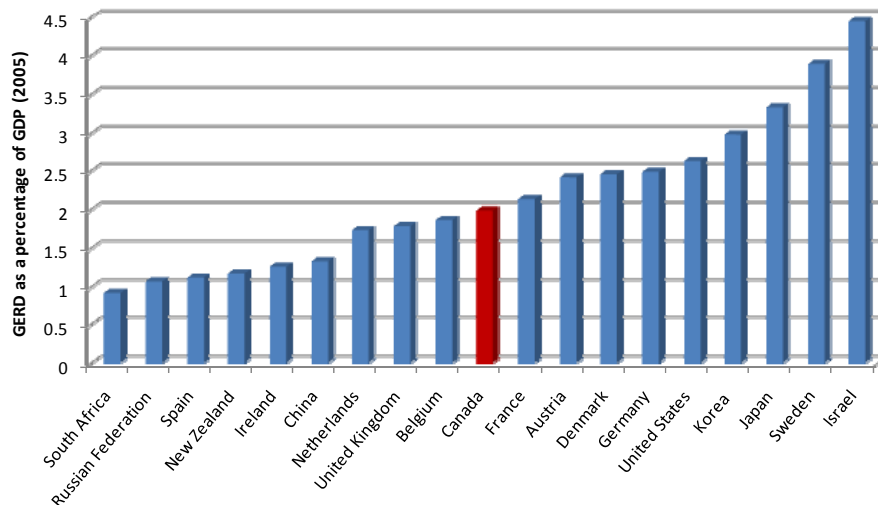
Figure 23. GERD as a percentage of GDP for selected countries

Source: OECD Fact book 2008, Expenditure on R&D (Organisation for Economic Co-Operation and Development 2008)

In this case, the figures used represent total funding for R&D (Gross Expenditure on R&D – GERD), not just that for health. In order to provide an overview without showing every country that the OECD has data on, select countries are shown (Organisation for Economic Co-Operation and Development 2008). These have been selected based on the total range of figures and comparability with Canada. For example, the UK, USA, Japan, Russian Federation, France and Germany are G8 countries, while the Netherlands, Sweden and Ireland are highlighted elsewhere in this assessment in terms of their ROI experience.

Outputs from health R&D

The OECD also collects data on outputs from R&D, including human resources for science and technology, patent outputs, innovation, and specific information on biotechnology.

In terms of R&D personnel, Canada ranks thirteenth in the OECD listing, behind Australia but ahead of Germany (Organisation for Economic Co-Operation and Development 2007). For patent outputs, Canada performs relatively well in terms of triadic patents (patents held with the EU, US and Japanese patent offices), ranking tenth in the list of countries but well behind the USA, Japan and Germany. However, when the triadic patent numbers are normalised for population (per million population), Canada's performance noticeably drops to eighteenth, well below the OECD average (Organisation for Economic Co-Operation and Development 2007).

For innovation, one of the measures used by the OECD is the percentage of firms that have a 'new to market' product. For large companies, Canada performs relatively poorly, lying around halfway through the list. But for Small Medium Enterprises (SMEs), Canada is second in this innovation measure, only behind Iceland (Organisation for Economic Co-Operation and Development 2007). Most of these measures are already collected in Canada, and published in the Treasury Board's annual performance report (Treasury Board of Canada Secretariat 2007), and accord to a plan to utilise Canadian research resources to their fullest with targets for performance based on OECD averages. None of this data is specific to health research however, only to R&D as a whole. OECD data on biotechnology outputs

provides a starting point for accessing information on outputs in health research but does not give a full picture.

Bibliometric analysis provides a relatively simple and useful tool for addressing the issue of outputs from Canadian R&D in health, since publications can be identified within specific research fields and as arising from specific locations. They are also instantly internationally comparable along a number of different axes. For example, health research publications can be compared in terms of pure volume of publications, volume of citations, citations per publication, or the percentage of the world's most highly cited papers. It is worth noting that bibliometric analysis brings with it two inherent time lags. Firstly, the time from funding to publication, and secondly from publication to citation. The first is typically the length of the funding received, around 3-4 years. The second is an ongoing window, but in citation analysis the use of a four year citation window is common, implying that research four years or older can be analysed (Nason, Grant et al. 2007)(Moed 2005)(Glanzel, Moed 2002).

Historically, Canada produced around 4% of the world's health related publications in the 1992-2001 period (Paraje, Sadana et al. 2005), some way behind the US (37%) and trailing the UK, Japan, Germany and France (8%-5%). Of the publications coming out of Canada the majority are in clinical medicine (Figure 24), with several other fields related to health contributing to the output of the country (Sciencewatch.com 2008a).
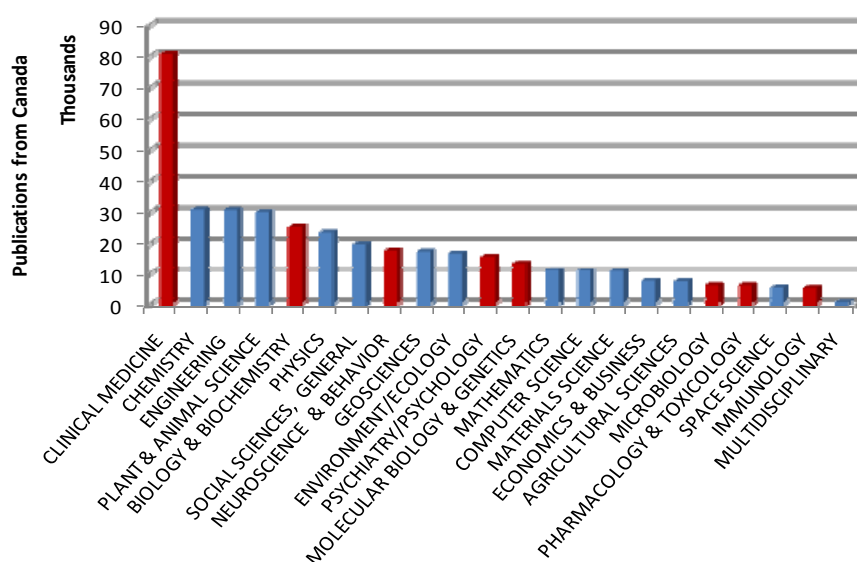

Figure 24. Publications from Canada 1998-2008 in the ISI database. Research fields related to health are shown in red.

Source: (Sciencewatch.com 2008a)

In order to judge the quality of these publications people often use citations as a discerning factor. The ISI (Institute for Scientific Information) database of publications and citations that holds this information on Canada identifies the total citations for each research field, again with clinical medicine coming out as top of Canadian research (Sciencewatch.com 2008a). However, with the publication and citation numbers we encounter one of the main problems of bibliometrics. Namely, that different research fields have different publication and citation behaviours (Moed 2005), so comparing publication or citation counts gives a false impression of the volume and quality of work going on within a discipline. Using citations per paper helps with this slightly since it normalises the citation measure to relate to the

volume of publications. However it still does not allow comparison across disciplines since it doesn't take into account the behaviour of the discipline as a whole. For that we need to use a relative citation impact that uses the average citations per paper in a particular field and then divides the citations per paper for the research performed in Canada in that field by the field average (thus suggesting whether the research is better than the world average (>1) or worse (<1)). Notably, all disciplines shown in Figure 25 are at or above the world average (≥1).
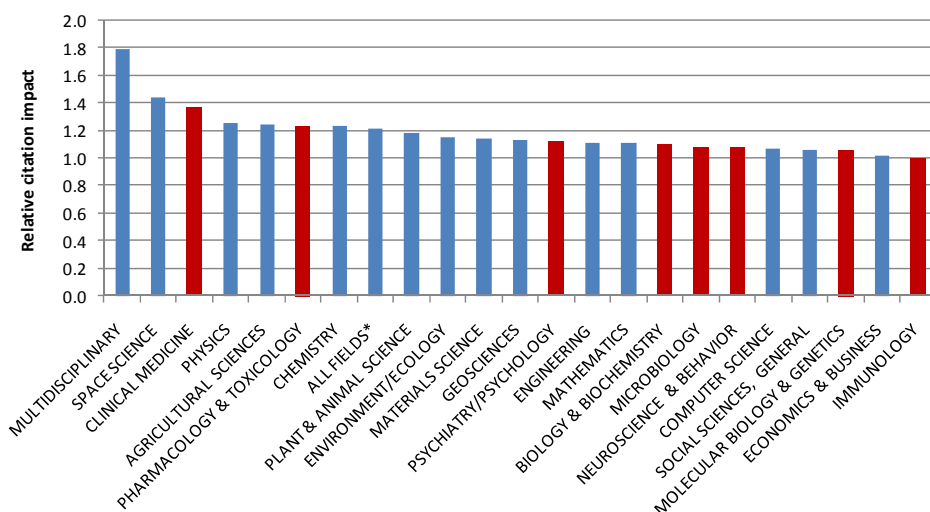
Figure 25. Relative citation impact for Canadian publications in the ESI database (1998-2008). Research fields related to health are shown in red.

Source: Data taken from Essential Science Indicators (subscription website) - {{637 Thomson Reuters 2008}}

By understanding Canada's share of publications in the world (Figure 26) we can identify disciplines producing a large proportion of publications. It should be noted this tells us nothing about the quality of the research in these fields, something Figure 25 shows using relative citation impact.
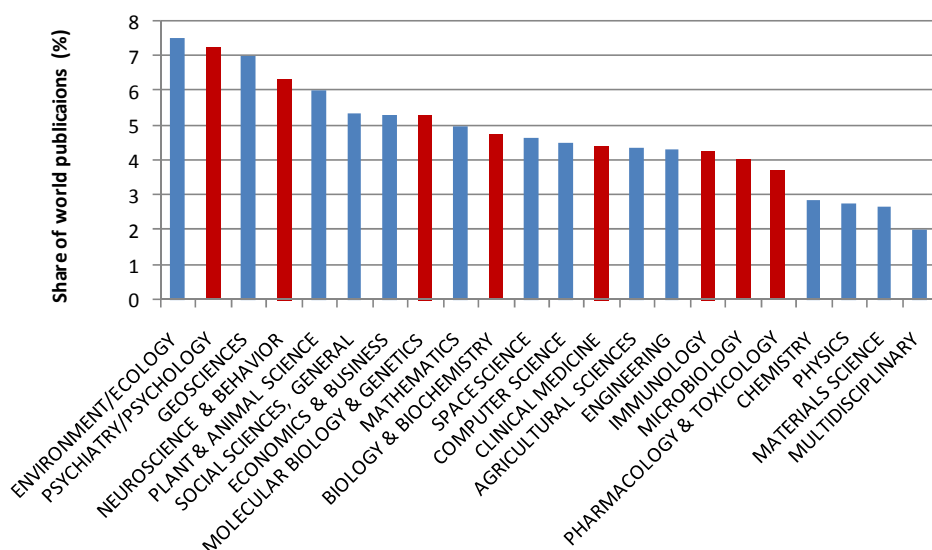
Figure 26. Percentage share of world publications held by Canada for different publication fields (1998-2008). Health related fields are shown in red.

We can relatively safely use citations per paper as a measure of research quality when considering a single discipline, such as clinical medicine (Figure 27), to identify where Canada sits in terms of producing high quality research – although there may also be issues over Anglophone bias in the citation of research (Moed 2005).
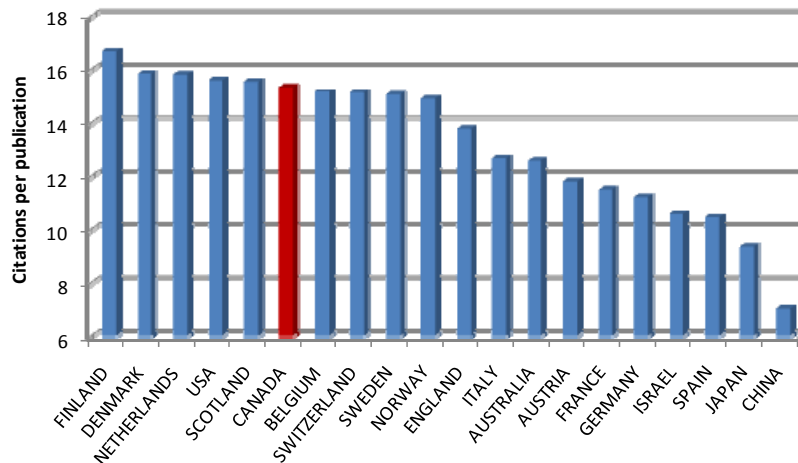
Figure 27. Citations per publication for clinical medicine research in the top 20 most cited countries (ISI database publications only; 1998-2008)

Source: (Sciencewatch.com 2008b)

In Canada, although there is no systematic collection of research output data specifically for health, there are pockets of analysis. Science-Metrix (a bibliometric analysis consultancy) has produced analyses of health research impacts for many provinces (Science-Metrix 2008), CIHR and the Heart and Stroke Foundation are currently involved in a project that will provide a bibliometric analysis of cardiovascular research in Canada against that in the UK and Australia (amongst other countries) (Heart and Stroke Foundation of Canada 2008), and in 2006 an analysis of a variety of biomedical research fields (Falagas, Michalopoulos et al. 2006) placed Canada first in the world for quality in biomedical research based on three research areas, a quality measure using volume of publications multiplied by journal impact factor for the research field, and then normalised using GDP to show what the authors call 'efficiency in research'. Although this is a very positive finding, it must be taken with some caution. The choice of biomedical research fields has no explanation as to why these three areas were selected so may not be representative; the quality measure relies on journal impact factors which do not reflect the quality of individual papers within them; and the use of GDP as a normalising factor can only be treated as a means of determining effectiveness if each country spent the same proportion of GDP on health research (something that Figure 23 shows us not to be the case). Other work that looks at the position of different nations in health research also places Canada high in the G8 nations for medical research citations (King 2004), but there is no study that compares health research as a whole for different countries on quality.

Conclusion

What we have seen in this chapter is that the structure of the health research system in Canada is a complex one with many players involved in the funding and production of health research, from government through to private organizations. Each of these groups understandably has their own

evaluation needs. This complexity is added to by the diversity of health research disciplines, generally considered in Canada to be classified within at least one of the four pillars of research. Understanding the priorities of funders of health research, particularly those sponsoring this assessment, is a requirement for producing a valid evaluation framework and metrics set. In order to link the inputs to the outputs and outcomes of health research in Canada, there is a need to understand what methodologies exist for effective and comprehensive evaluation; methods that can cope with the complexity of the system. Appendix C goes into detail on the frameworks and evaluation methods available to investigate the impacts of health research.

References

BEARING POINT, 2004. *Interim Evaluation of Genome Canada.*

BEAUDET, A., 2007. Return on Investments in Health Research: Defining the Best Metrics - A Provincial Perspective. Presentation to CAHS Third Annual Meeting edn. Montreal, Quebec.

BERNSTEIN A, HICKS V, BORBEY P, CAMPBELL T, 2006. *A framework to measure the impact of investments in health research.* OECD Blue Sky Forum II.

BUXTON, M.J. and HANNEY, S.R., 1996. How can payback from health services research be assessed? *Journal of Health Services Research Policy,* **1**(1), pp. 35-43.

CANADA EXCELLENCE RESEARCH CHAIRS, , canada excellence research chairs frequently asked questions. Available: http://www.cerc.gc.ca/cpfa-pcfa-eng.shtml [10/09, .

CANADA FOUNDATION FOR INNOVATION, 2006. *Solid Foundations: Leading Innovations through Research Infrastructure - Annual Report 2005-2006.*

CANADA FOUNDATION FOR INNOVATION, 2005. *Results through Innovation, 2004-2005 Annual Report.*

CANADA FOUNDATION FOR INNOVATION, , funded projected database. Available: http://www2.innovation.ca/pls/fci/fcienrep.base.

CANADA RESEARCH CHAIRS, 2/21/2008, 2008-last update, about us: thinking ahead. Available: http://www.chairs.gc.ca/web/home_e.asp [10/08, 2008].

CANADA RESEARCH CHAIRS, 2006. *Integrated Results-based Management and Accountability Framework and Risk-Based Audit Framework.*

CANADIAN HEALTH SERVICES RESEARCH FOUNDATION, n.d.-last update, about CHSRF [Homepage of CHSRF], [Online]. Available: http://www.chsrf.ca/about/index_e.php [August 27th, 2008].

CANADIAN INSTITUTE FOR HEALTH INFORMATION, 2007. *National Health Expenditure Trends: 1975-2007.* Ottawa, Ontario: CIHI.

CANADIAN INSTITUTES FOR HEALTH RESEARCH, 2005. *Developing a CIHR Framework to Measure the Impact of Health Research.* Ottawa, Ontario: CIHR.

CANADIAN INSTITUTES OF HEALTH RESEARCH, 2008-last update, current budget. Available: http://www.cihr-irsc.gc.ca/e/22953.html.

CANADIAN INSTITUTES OF HEALTH RESEARCH, September 17th 2007, 2007-last update, an overview of CIHR [Homepage of Camadian Institutes of Health Research], [Online]. Available: http://www.cihr-irsc.gc.ca/e/30240.html#slide1_e [June 26th, 2008].

CIHR, 2008. Payback categories. Ottawa, Canada: CIHR.

FALAGAS, M.E., MICHALOPOULOS, A.S., BLIZIOTIS, I.A. and SOTERIADES, E.S., 2006. A bibliometric analysis by geographic area of published research in several biomedical fields, 1995–2003. *CMAJ,* **175**(11), pp. 1389-1390.

FAST, E., 2007. *Mobilizing Science and Technology: The New Federal Strategy.* PRB 07-32E. Ottawa, Ontario: Library of Parliament.

GENOME CANADA, 2008a-last update, genome canada - about. Available: http://www.genomecanada.ca/en/about/.

GENOME CANADA, 2008b. *Science Leads Us: 2007-08 Annual Report.*

GENOME CANADA, 2007. *Performance, Audit and Evaluation Strategy.*

GLANZEL, W. and MOED, H.F., 2002. Journal impact measures in bibliometric research. *Scientometrics,* **53**(2), pp. 171-193.

GOVERNMENT OF CANADA, n.d. *Canadian Medical Devices Industry.*

GOVERNMENT OF CANADA, 9/19/2008, 2008a-last update, indirect costs program. Available: http://www.indirectcosts.gc.ca/about/index_e.asp [10/14, 2008].

GOVERNMENT OF CANADA, 2008b. *Science and Technology Data 2006.* Ottawa, Ontario: Government of Canada.

GOVERNMENT OF CANADA, 2007-last update, doing business with canada: health industries. Available: http://www.canadainternational.gc.ca/dbc/Medical-Devices-Sector-Canada-en.aspx.

HEART AND STROKE FOUNDATION OF CANADA, 2008-last update, its payback time: new international study to assess impact of heart and stroke research [Homepage of HSFC], [Online]. Available: http://www.heartandstroke.com/site/apps/nlnet/content2.aspx?c=ikIQLcMWJtE&b=3485819&ct=4512407 [July 24th, 2008].

HEART AND STROKE FOUNDATION OF CANADA, 2006. *Research: Heart and Stroke Foundation of Canada Research Report - 2005/2006.* Ottawa, Ontario: Heart and Stroke Foundation of Canada.

HM TREASURY, CABINET OFFICE, NATIONAL AUDIT OFFICE, AUDIT COMMISSION and OFFICE FOR NATIONAL STATISTICS, 2001. *Choosing the Right FABRIC: A Framework for Performance Information.* London, UK: HM Stationary Office.

HONOURABLE JAMES M. FLAHERTY, MINISTER OF FINANCE, 2007. The Budget Plan 2007: Aspire to a Stronger, Safer, Better Canada.

INDUSTRY CANADA, 2008a-last update, canada's biotechnology industry. Available: http://www.ic.gc.ca/epic/site/lsg-pdsv.nsf/en/h_hn00079e.html.

INDUSTRY CANADA, 2008b-last update, canadian medical device industry. Available: http://www.ic.gc.ca/epic/site/md-am.nsf/en/hi00083e.html.

KING, D.A., 2004. The scientific impact of nations. *Nature,* **430**(6997), pp. 311-316.

MADORE, O., NORRIS, S., 2006. *Federal funding for health research.* PRB 06-27E. Ottawa: Library of Parliament.

MCGILL UNIVERSITY, Spring 2008. *Making History: The Campaign McGill Newsletter.*

MOED, H.F., 2005. Citation Analysis in Research Evaluation. Dordrecht, Netherlands: Springer.

NASON, E., 2008. *Health and Medical Research in Canada.* RAND Europe.

NASON, E., GRANT, J. and VAN LEEUWEN, T.N., 2007. *Bibliometric analysis of highly cited publications of health research in England, 1997-2003:Theme specific HCPs in England.* WR-509-DH. Cambridge, UK: RAND Europe.

NATIONAL CANCER INSTITUTE OF CANADA, 2007. *Financial Statements of National Cancer Institute of Canada: Year ended March 31st 2007.* Toronto, Ontario: KPMG.

NATIONAL INSTITUTES OF HEALTH, 2006-last update, NIH extramural awards by state and foreign sites. Available: http://report.nih.gov/award/state/state06.cfm.

NATIONAL RESEARCH COUNCIL OF CANADA, 2008-last update, NRC biotechnology research institute. Available: http://www.irb-bri.cnrc-nrc.gc.ca/home/index_e.html.

NETWORKS OF CENTRES OF EXCELLENCE, 2008-last update, networks of centres of excellence: 'forging innovative networks'. Available: [http://www.nce.gc.ca/comp/CECR/NCENewProgramsJuly2007.pdfpublications.

NETWORKS OF CENTRES OF EXCELLENCE, June 2002, 2002a-last update, results-based management and accountability framework. Available: http://www.nce.gc.ca/pubs/reports/2002/rmaf/rmaf_e.htm.

NETWORKS OF CENTRES OF EXCELLENCE, September 2002, 2002b-last update, risk-based audit framework. Available: http://www.nce.gc.ca/pubs/reports/2002/rbaf/rbaf_e.htm.

NETWORKS OF CENTRES OF EXCELLENCE, , annual report 06/07. Available: http://www.nce.gc.ca/annualreport2006_2007/eng/2_2/2_2_6-eng.asp#pie.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, 2008. *OECD Factbook 2008: Economic, Environmental and Social Statistics.* Paris, France: OECD.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, 2007. *OECD Science, Technology and Industry Scoreboard 2007: Innovation and Performance in the Global Economy.* Paris, France: OECD.

PARAJE, G., SADANA, R. and KARAM, G., 2005. Increasing International Gaps in Health-Related Publications. *Science,* **308**, pp. 959-960.

SCIENCE, INNOVATION AND ELECTRONIC INFORMATION DIVISION, 2008. *Science Statistics: Estimates of Total Spending on Research and Development in the Health Field in Canada, 1996 to 2007.* July 2008 edition. Ottawa, Ontario: Statistics Canada.

SCIENCE-METRIX, n.d. 2008-last update, reports [Homepage of Science-Metrix], [Online]. Available: http://www.science-metrix.com/eng/reports_2008_t.htm [August 12th, 2008].

SCIENCEWATCH.COM, n.d. 2008a-last update, country profiles 2008: canada. Available: http://sciencewatch.com/dr/cou/2008/08junCanada/ [August 12th, 2008].

SCIENCEWATCH.COM, n.d. 2008b-last update, country profiles 2008: top 20 countries in clinical medicine. Available: http://sciencewatch.com/dr/cou/2008/08aug20CLI/ [August 12th, 2008].

STATISTICS CANADA, 2008a-last update, the daily: research and development in the health field. Available: http://www.statcan.ca/Daily/English/080722/d080722d.htm.

STATISTICS CANADA, May 30th, 2008b-last update, domestic spending on research and development (GERD) [Homepage of Statistics Canada], [Online]. Available: http://www40.statcan.ca/l01/cst01/scte03.htm [September 8th, 2008].

STATISTICS CANADA, May 30th, 2008c-last update, gross domestic product, expenditure-based [Homepage of Statistics Canada], [Online]. Available: http://www40.statcan.ca/l01/cst01/econ04.htm [August 10th, 2008].

STATISTICS CANADA, September 5th, 2008d-last update, research and development performed by the business enterprise sector (all industries) [Homepage of Statistics Canada], [Online]. Available: http://www40.statcan.ca/l01/cst01/econ151a.htm [September 8th, 2008].

STATISTICS CANADA, 2007. *Federal Scientific Activities 2006/2007.* 88-204-XIE.

STATISTICS CANADA, 2005. *Cornerstones of Community: Highlights of the National Survey of Nonprofit and Voluntary Organizations.* 61-533-XIE.

THE CONFERENCE BOARD OF CANADA, 2007. *How Canada Performs: A Report Card on Canada.* Ottawa, Ontario: The Conference Board of Canada.

THE CONFERENCE BOARD OF CANADA, 2004. *Exploring Canada's Innovation Character: Benchmarking Against Global Best.* Ottawa, Ontario: The Conference Board of Canada.

TREASURY BOARD OF CANADA SECRETARIAT, 2007. *Canada's Performance Report 2006-07: The Government of Canada's Contribution.* Ottawa, Ontario: Treasury Board of Canada Secretariat.

UNIVERSITY OF TORONTO, , 1997-2004 fundraising campaign. Available: http://www.artsci.utoronto.ca/main/givingtoartsandscience/friends/campaignroundup [10/10, 2008].

WAHL, A., 2008. The power of innovation. *Canadian Business,* **March 17th**.

Appendix C: Evaluation frameworks and methods

Canada's health research system brings together funders and researchers from provincial and federal government, higher education, business and non-profit sectors. With this many different stakeholders, it is unsurprising that the impacts of research (both intended and unintended) are wide-ranging. An organized framework is required to adequately capture and understand the variety of impacts arising from research.

## Evaluation Frameworks: What is available for use?

In this section we will identify different evaluation frameworks and the categories they fall into. We cover the way these frameworks work and will discuss uses of particular frameworks in later sections. In this section we will cover the Payback Framework, the Walt and Gilson Analytical Model, Research Impact Framework, Research Utilization Ladder, the Lavis Decision Making Impact Model, the Weiss Logic Model Approach, HTA Organization Assessment Framework, Societal Impact Framework, and the Balanced Scorecard.

The Payback Framework

The Payback Framework is built around a logic model that demonstrates where benefits from research are likely to appear and how they can best be assessed. The Framework also contains a multi-dimensional categorization of benefits arising from research. Developed originally for assessing the payback, or benefits, from health services research (Buxton, Hanney 1994, Buxton, Hanney 1996) the Payback Framework has been applied to various types of health research from HTA (Hanney, Buxton et al. 2007), through clinical research (Hanney, Frame et al. 2003) to basic biomedical research (Hanney, Mugford et al. 2005).

The logic model for the Payback Framework (Figure 28) provides a structure for analyzing the progress of a research idea from inception (Stage 0) through the research process (Stage 2) into dissemination (Interface B) and on towards its impact on people and society (Stage 6). In this sense, the logic model is a tool to trace the progress of knowledge and its subsequent utilization, thereby helping to facilitate analysis and consistency in research techniques for data gathering. It does this by providing a common structure for all evaluations, thereby ensuring cognate information for each study is recorded in the same place. The model is not meant to imply that the research process itself is linear, rather, the presence of multiple feedback loops allows for research to influence 'earlier' stages in the process. The logic model is compatible with flows of knowledge to potential users and can demonstrate both slow and rapid diffusions of knowledge into the stock of knowledge.
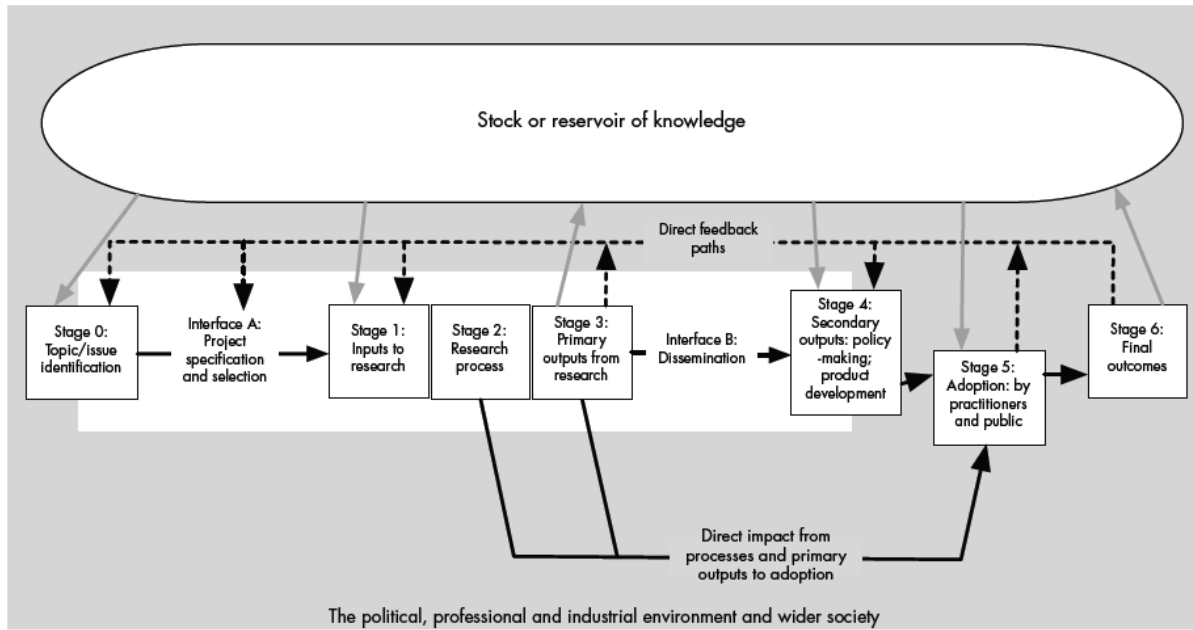
Figure 28. The Payback Framework logic model of research progress to impacts

Source: (Wooding 2006) *Reproduced with kind permission from the authors*

The categorization of impacts arising from research forms the second part of the Payback Framework with five areas of impact to categorize data (). Data can be collected using a wide range of methods in the Payback Framework, since the logic model and categories do not prescribe what impacts should be, but try to identify where they are. While it is not completely possible to tie the categories of benefits to specific stages of the model, it is possible to identify broad correlations: the knowledge production and research targeting and capacity building categories together are generally the primary outputs from research; the informing policy and product development category relates to the secondary outputs; and the categories for health and health sector benefits and broader economic benefits, respectively, are generally the final outcomes.

Box 2. Categories of impact used in the Payback Framework, with examples of data collected

Source: (Nason, Janta et al. 2008)

Knowledge production

Journal articles, conference presentations, books, book chapters, research reports

Research targeting and capacity building

Better targeting of future research, development of research skills, personnel and overall research capacity, staff development and educational benefits

Informing policy and product development

Improved information bases for political and executive decisions, developing pharmaceutical products and therapeutic techniques

Health and health sector benefits

Improved health, cost reduction in delivering existing services, qualitative improvements in the process of delivery, improved equity in service delivery

Broader economic and social benefits

Wider economic benefits from commercial exploitation of innovations arising from R&D, economic benefits from a healthy workforce and reduction in working days lost

---

Attempting to create a comprehensive framework for assessment of impacts has received some criticism, with some claiming it is unrealistic to be so comprehensive (Wooding, Nason et al. 2007). However, despite not always identifying final outcomes (Buxton, Hanney et al. 2000, Wooding, Hanney et al. 2004) the use of the Payback Framework has facilitated identification of some outcomes. The Payback Framework has also been accused of being resource intensive; however this assumption is mainly attributable to the use of case-study based evaluations using the framework. Since the Payback Framework is not prescriptive of the data collection methods it uses, payback studies can be performed using only surveys (Hanney, Buxton et al. 2007, Wooding 2008b), making them considerably less resource intensive.

The Payback Framework is widely used, with evaluations in multiple countries including Canada, Ireland, the Netherlands and Hong Kong (Nason, Janta et al. 2008, Canadian Institutes for Health Research 2005, Oortwijn, Hanney et al. 2008, Kwan, Johnston et al. 2007). It has also been used by international organizations, government, provincial and charity research funders (Hanney, Buxton et al. 2007, Buxton, Schneider 1999, Wooding, Hanney et al. 2005, Buxton, Hanney et al. 2004).

The Walt and Gilson Analytical Model

Developed primarily to address the issue of informing health policy, this model is a simplified view of the world in which policy making occurs. It places policy making in a dynamic relationship between four factors: the context of the world, the content of the policies, the process of policymaking itself and the actors involved in making policy (Walt, Gilson 1994). Actors are placed at the centre of the relationship since that is the group most likely to influence the other three aspects (Figure 29).
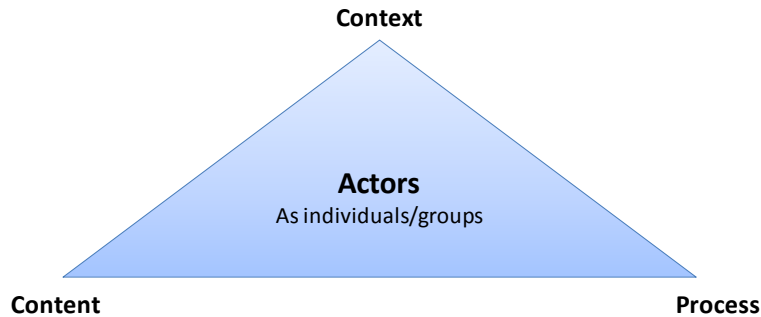
Figure 29. The Walt and Gilson Analytical Model (Walt, Gilson 1994)

Reproduced with kind permission from the authors

In terms of health R&D this model provides an important lesson. Namely, that the content of health policies is not the only aspect affecting the production of policy. From an evaluation point of view, this model can be utilized to investigate the ways that health research has gone on to influence policy, in light of the many competing factors informing policy. Although not widespread in evaluating R&D impact as a whole, this framework has been used to understand the way research informs health policy (Trostle, Bronfmann et al. 1999).

The Research Impact Framework

More recently, Gill Walt has been involved in the development of a framework more directed towards health R&D, which aims to provide a simple format for researchers themselves to use to identify and report their impacts (Kuruvilla, Mays et al. 2006). This framework builds around four broad areas of research impact: Research-related impacts; Policy impacts; Service impacts; and Societal impacts. Within each of these categories lie subcategories of impact (Box 3).

Box 3. The research impact framework – four categories of impact (Kuruvilla, Mays et al. 2006)

Research-related impacts: Type of problem/knowledge**;** Research methods**;** Publications and papers**;** Products, patents and translatability potential**;** Research networks**;** Leadership and awards**;** Research management**;** Communication.

Policy impacts: Level of policy-making; Type of policy; Nature of policy impact; Policy networks; Political capital.

Service impacts: Type of services: health/inter-sectoral**;** Evidence-based practice**;** Quality of care**;** Information systems**;** Services management**;** Cost-containment and cost-effectiveness.

Societal impacts: Knowledge, attitudes and behaviour; Health literacy; Health status; Equity and human rights; Macroeconomic/related to the economy; Social capital and empowerment; Culture and art; Sustainable development outcomes.

This framework has been used to evaluate the work of researchers based at London's School of Hygiene and Tropical Medicine, asking researchers to collect information in all the categories and subcategories of the model that represented their outcomes (Kuruvilla, Mays et al. 2006). One criticism of this framework would be that, despite providing categories in which to capture impacts, it does not provide any representation of the way in which impacts arise, making it a useful audit tool, but not so useful for organizational learning.

The Research Utilization Ladder

Another approach to understanding the impacts of health research has been to investigate the ways in which research progresses towards its application by practitioners and policy makers. In the Research Utilization Ladder Model, there are six distinct steps in the process from producing a piece of research to it being used by a practitioner or policy maker (Landry, Amara et al. 2001). These six steps are shown in Table 1 and represent the diminishing control of the researcher themselves as research findings move 'up the ladder'.

Table 1. The research utilization ladder (Landry, Amara et al. 2001)

| Stage 1	**Transmission:** transmitted research results to practitioners and professionals.
Stage 2	**Cognition**: research reports read and understood by practitioners and professionals.
Stage 3	**Reference:** work cited as a reference in reports, studies, and strategies of action elaborated by practitioners and professionals.
Stage 4	**Effort:** efforts made to adopt research results by practitioners and professionals.
Stage 5	**Influence:** research results influenced choices and decisions of practitioners and professionals.
Stage 6	**Application:** research results gave rise to applications and extension by practitioners and professionals.

The Research Utilization Ladder is one way of visualizing the concept of research utilization, one aspect of health R&D. As an evaluative tool for R&D as a whole, this is an incomplete model since it can only examine knowledge translation, and not research quality, integrity or efficiency. It also does not describe any final outcomes from the research once it has been adopted by practice or policy. In of itself, this is a useful tool for examining research uptake and has been used for understanding research utilization by nurses (Profetto-McGrath, Hesketh et al. 2003). For the purposes of R&D evaluation this would not cover enough aspects of R&D to be considered useful, although it could feed into thinking on maximizing knowledge translation from R&D.

The Lavis Decision Making Impact Model

Impacting on the decision making process of any individual or organization is the basic tenet of this model (Lavis, Ross et al. 2003). By investigating the way that research influences decision making, it can investigate comparative levels of impact. The model is built around four questions, each with their own sub-questions. Firstly, who are the target audiences for the research? Within this category, the model identifies researchers, the general public, patients, clinicians, care managers, R&D officers and policy makers. Secondly, how can we measure impact? This is done by using three categories of promoting research: user-pull (decision-makers are the ones seeking research); producer-push (researchers actively disseminating results); and exchange measures (where both parties are involved actively). For each of these categories there are a number of metrics identified. Thirdly, what are the metrics most appropriate for the category of knowledge exchange used in the example of research being evaluated? Finally, what are the most appropriate measures considering the constraints of the evaluation itself (budget constraints, timescale of the evaluation, etc.)? These are shown in Figure 30.

Since this model focuses on the way decision making is influenced it can address any kind of impact that involves a decision being made. This could be a decision to cite a publication (implying citations can be used as metric), through to a decision by an individual to not smoke (a final health outcome). This is a great strength of this model. Its main weakness lies in the need to identify the target audience at the

beginning of any evaluation, making the model likely to miss unintended outcomes (be they positive or negative) in any other audience.
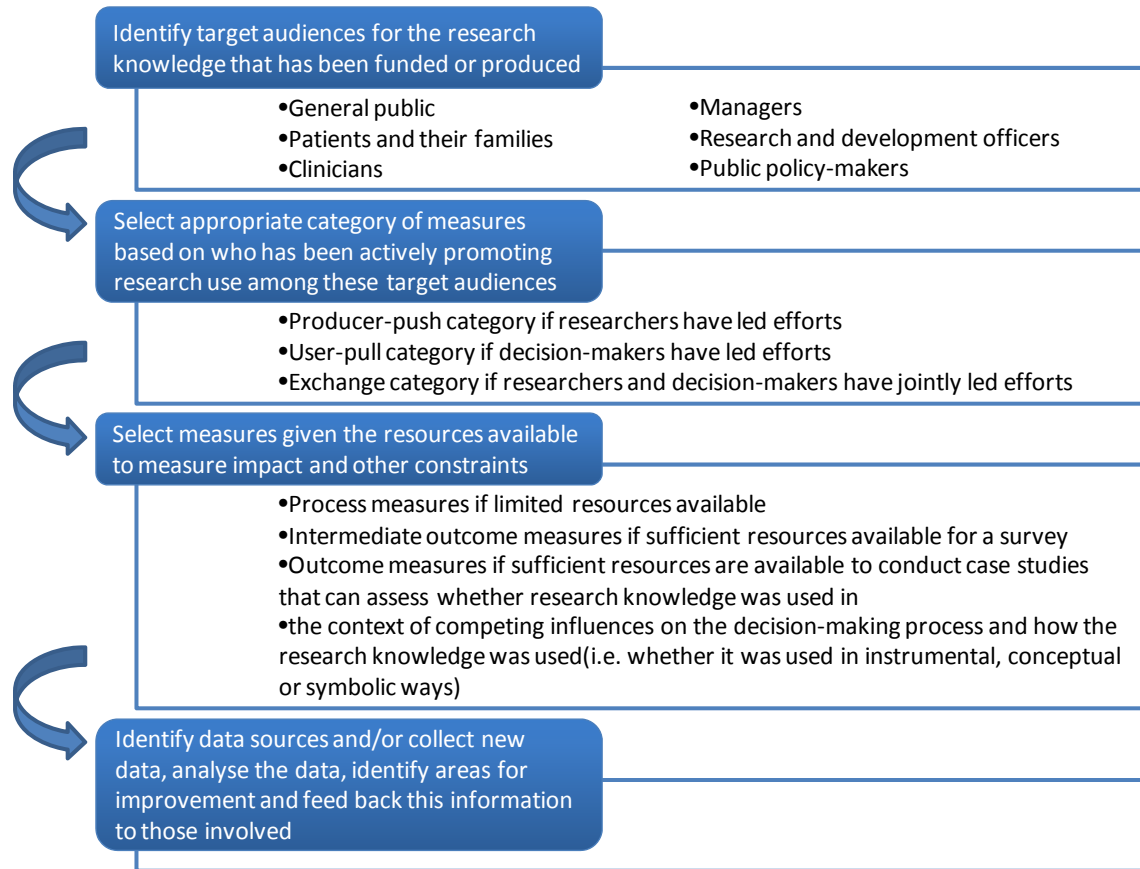

Figure 30. The four stages of the decision making impact model

Source: Lavis, Ross et al. 2003

The Weiss Logic Model Approach

In an attempt to understand the outcomes from medical research, rather than the traditional evaluation view of identifying outputs from research, this evaluation approach builds on the logic modelling of research outputs designed by the United Way in the US in 1996 (Weiss 2007). In the original logic model, the assessment of inputs, process and outputs was explicit (as resources, activities and products). However, the outcomes were simply seen as the 'benefits or changes in a population of interest'. The Weiss model takes the concept of outcomes from medical research and splits them into three sub-categories of outcome: initial, intermediate and long-term. Initial outcomes are simply a raising of awareness of medical research in the decision making community; intermediate outcomes are any changes in practice that arise from that increased awareness; long-term outcomes are the accompanying changes to the health of patients.

This model allows linkage of the clinical outcomes that arise from treating patients and the research outputs from medical research. Weiss describes this as bridging the 'efficacy-effectiveness' gap (Figure 31), a way to understand the effectiveness of the research in fulfilling the final aim of medical research – improved clinical outcomes.
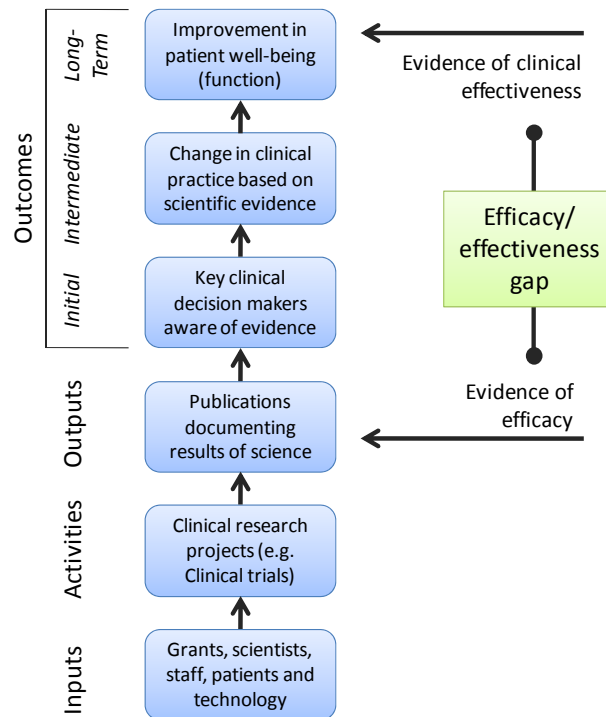
Figure 31. Logic model approach to evaluating medical research (Weiss 2007)

Reproduced with kind permission from the authors

As a new model, this logic model approach has not been tested with an evaluation of any specific medical research, however, this methodology has been proposed as a way to understand the return on investment in anti-doping research for sports (Lippi, Franchini et al. 2008). On a theoretical level, Weiss concedes that although it is relatively easy to get good data on the outputs of research, trying to get reliable indicators that can be linked to the research for outcomes is more difficult. From a health research point of view, this model is also limited since it only covers medical research.

HTA Organization Assessment Framework

Recent work in Montreal has drawn upon a wealth of evaluation work on Health Technology Assessment (HTA) organizations and the extensive literature on indicators and metrics for HTA to address the issue of impact assessment for HTA organizations (Lafortune, Farand et al. 2008). The authors base their model on Parsons' social action theory. The model builds around four functions an organization needs to perform well in order to succeed:

- **Goal attainment**

    The effectiveness and efficiency with which the organization achieves its mission. Effectiveness is measured by the ability to impact on decision making; the way health and clinical services are organized and managed; on health and society more generally; and on further research (particularly in HTA). Efficiency is considered to be performing HTA functions with the best (most prudent) use of resources.

- **Production**

    Producing outputs is the main aim of an HTA organization. Measures of production in this model are: volume of output; productivity of output – comparing the output volume to the

resources used in achieving that output; coordination mechanisms to look at the way processes are managed in the production of outputs; and the quality of outputs based on their accessibility for users, how comprehensive the output is, and the technical quality of the HTA report itself.

- **Adaptation to the environment**

    By being more adaptable, an HTA organization can better produce outputs that fit the environment in which they work. First, by having a capacity to attract and maintain resources (be those human, financial, technical or other). Second, by maintaining an ability to mobilize external support through improved visibility and credibility. Third, by showing an ability to respond to changing needs when required. Fourth, by innovating and altering the HTA knowledge field and learning from organizational actions.

- **Culture and values maintenance**

    The model assumes that there are three principle values around which organizations can build culture: independence in action, transparency of process, and accountability to stakeholders. Added to these principle values are a number of factors that create a successful organizational climate which includes leadership qualities, teamwork, communication skills and motivation.

Between each of these four categories, there are linkages that show how different aspects align with one another to improve the overall impact of the organization (Figure 32).
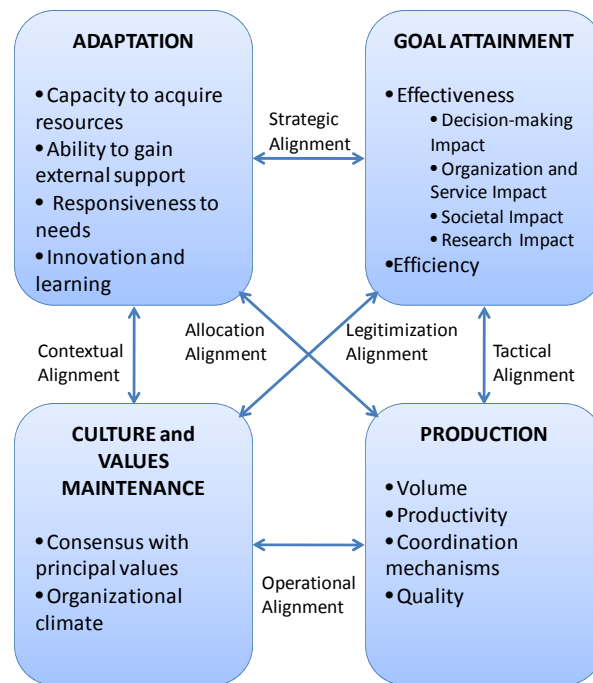
Figure 32. HTA organization assessment framework (Lafortune, Farand et al. 2008)
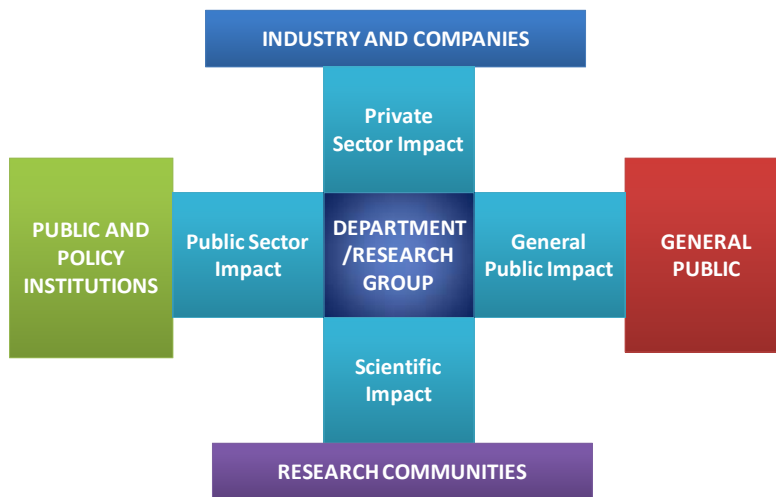
Reproduced with kind permission from the authors

Since this work is very new (only published in January 2008) it has not had time to be implemented anywhere. However, the authors have performed an extensive review of evaluations of HTA organizations to inform this publication so it is built upon real-life evidence. However, since this has all been bought together it remains to be seen how this would be implemented as a single framework.

Also, this framework looks exclusively at HTA, which although an important aspect of health research, is only a very small fraction of the totality.

The Societal Impact Framework

The Societal Impact Framework is built around the 'communication metaphor' (van Ark 2007) which states that "evaluating the outcomes of R&D is considered as the valuation of the communication of research groups with relevant surroundings." This links to societal impact, since the communication with different societal communities can be assessed and valued. The framework identifies four different societal groups that can be communicated to produce an impact: Industry; the general public; the scientific community; and public and policy institutions (Figure 33).

Figure 33. The four societal communities and the impacts of research by a research group (van Ark 2007)

Reproduced with kind permission from the authors

In order to measure impact on these communities, the societal impact model uses four categories of communication: knowledge products; knowledge exchange and esteem; knowledge use; and attractiveness. Knowledge products represent the direct outputs of research (publications, patents, products, etc.); knowledge exchange includes presentations, consultancy and public lectures; knowledge use represents the use of research findings and is measured using citations, use of products, etc.; attractiveness is essentially the way that research brings in further funding (from whichever source that may be).

The distinction of identifying who research is aimed at is a useful one in trying to understand the processes around research translation, but this model does not go on to try and evaluate the impact of research on final outcomes such as the economic benefits to society itself or the health benefits arising from research.

Research Embedment and Performance Profiles (REPP), a modification of the societal impacts method, captures impacts from in five domains (Box 4).

Box 4. The five domains of the REPP (Council for Medical Sciences 2002)

Science and certified knowledge: the production of knowledge claims for validation by the scientific community (peers);

Education and training: the training of researchers and the generation of skills (embodied or tacit knowledge);

Innovation and professionalism: the production of knowledge with a view to gaining a competitive advantage;

Public policy: the production of knowledge and skills for policy or social purposes;

Collaboration and visibility: 'internal' orientation and performance in the contribution to the research organization's goals as compared with orientation towards other national and international institutions.

By scoring these five domains using a variety of metrics (such as publications produced and presentations given) one can plot a profile on a spider-plot to show where a research organization has strengths and weaknesses (SciQuest n.d.). As with the first version of the societal impacts framework, this method suffers from not being able to incorporate any final outcomes from research, while also being restricted to investigating research institutes or universities in order to be able to create comparable REPPs.

The Balanced Scorecard

The Balanced Scorecard (BSC) (Kaplan, Norton 1992) is a relatively comprehensive approach to measuring performance and driving organizational strategy. The BSC attempts to overcome a narrow focus on financial performance (or socio-economic outcomes for public entities) by incorporating other perspectives on the performance of an organization. The BSC balances out the focus on simple financial performance by incorporating three other performance aspects of the organization: the customer; business process; and learning and growth (Figure 34). Each of these four areas can be scored for objectives, measures, targets and initiatives using specific metrics in each area that can be designed for the organization being evaluated.
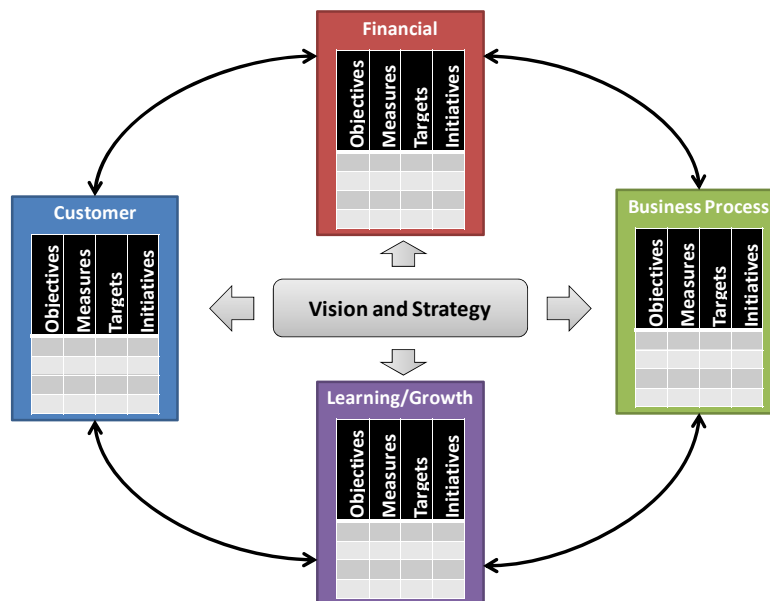
Figure 34. The Balanced Scorecard

Kaplan and Norton argue that the BSC is not simply a performance measurement tool, but a strategic management system (Kaplan, Norton 1996). In developing metrics for all four areas, the BSC forces managers to focus on those aspects of the organization that are most important to future success. In this manner, the BSC helps managers translate strategy into operational goals, plan accordingly, measure performance in all areas and adjust strategy accordingly (Kaplan, Norton 1996). This methodology has been hugely successful and popular in business, with major companies such as IBM and Ford using variations of it in their management strategies (Balanced Scorecard Institute 2007), although the BSC has also generated interest in the public sector (Modell 2004). With regard to R&D, the BSC has found limited application, focusing on the later stages of R&D, knowledge transfer to commercialization, and financial benefit (Bremser, Barsky 2004, Eliat, Golany et al. 2006).

There is no reason to think the BSC is analytically unsuited to R&D (Osama 2006), and it has been suggested that it could be utilized with modifications of the four categories: for example, suggesting that innovativeness, speed of innovation, and cost savings from innovation are more appropriate than research cost savings and research productivity (Jordan, Mote et al. 2006). In Canada, Health Canada considered the use of the BSC during their development of an evaluation framework for electronic health records (Neville, Gates et al. 2003) and the Ontario University Health Network use a modified BSC to evaluate their performance for health and research activities (University Health Network 2008a). The major difficulty of using the BSC is that it is not best suited to organizational learning; where logic model based frameworks have an advantage.

Overview of frameworks

Each framework mentioned above has merits and drawbacks, as discussed for each framework. The main drawback with some of these models is that they are not applicable to health research in general, and though they may well guide detailed evaluations of the sections of health research they do address, they are not comparable to other sectors of research. Of those frameworks that are applicable across health research (the payback model, the research impact framework, the research utilization ladder, the decision making impact model, the societal impact framework, and the balanced scorecard), there are other reasons that make some unsuitable for addressing the impacts of all health research in Canada. The research utilization ladder is designed to look specifically at knowledge translation. The research impact framework and the balanced scorecard do not identify the processes involved in achieving impacts, making them useful for audit but less so for organizational learning. The decision making impacts model requires identification of the intended audience for the research, missing out on unintended impacts and making it difficult to compare research for different intended audiences. The Societal impact model has issues with identifying the final outcomes of health research, whilst the payback model as a comprehensive tool could lack analytical power. In the following section we identify how some of these frameworks have been implemented in Canada and around the world.

Practical application: Examples of frameworks in use in Canada

It would be foolhardy to attempt to understand the return on investment in health research in Canada without first understanding how this is currently addressed by the various stakeholders in Canadian health research. Since the remit of this project is wide enough to cover every type of health research performed in the country, it also must cover every type of health research funding agency in the country. This is particularly important because it addresses a key issue. The groups most interested in understanding a return on investment are those that are most directly affected by it, the funders

themselves, since they are the ones who have a mission to be accountable to in terms of impacts (effectiveness), and financial backing that they must show returns on (efficiency).

At all levels there is some sort of evaluation occurring, with different organizations using different evaluation techniques; many producing ad hoc evaluations of particular funding streams. Some organizations have built in evaluation frameworks or systems that allow them to assess their impact at an organizational level, and it is these that are particularly interesting to us within this report. Ad hoc evaluations provide information on the sorts of impacts that can occur from health research in Canada, but do not give an idea of how to collect data in an ongoing fashion, something that is needed for developing indicators for the health research system as a whole.

In this section we cover the different levels at which research is funded in Canada, citing examples of the evaluation techniques used at each level. The highest level of any evaluation of health R&D is at the international level, comparing the performance of the whole system against health R&D in other countries. Below that level, federal research funders need to understand the outputs and impacts of their funded research; whilst provincial funders need to assess impacts and compare themselves to other provinces. In general research institutes and universities rarely set up frameworks for evaluation, preferring to perform one of evaluations of programs or utilise others' frameworks (Association of Universities and Colleges of Canada 2006, Brimacombe, Gruenwoldt November 2007). There are exceptions to this however, with university networks coming together to evaluate the impacts from their health providers (University Health Network 2008a).

Other funders also evaluate their impacts, with different ends in mind and different stakeholders to be accountable to. We assign these to three main groups: Industry funders of R&D (often performing the R&D in house); independent foundations (often using endowment funds or funding research students); and charity funders of research. Each of these groups also has evaluation systems in place, and an example from each is addressed.

Statistics Canada: R&D Competitiveness

At an international level, there is already data collected and comparisons made as to the state of R&D in Canada. This covers both international competitiveness through the comparison of OECD collected data and intra-national comparisons of provinces, research funders and sectors carrying out research. These comparisons, discussed in the previous chapter on the position of Canadian health research in the world, do not truly constitute an evaluation system however, since they are distinct measures that do not conform to any particular goal or mission. The section on the position that Canada has in the world for health R&D is a prime example of the sorts of international comparisons that can be made on a regular basis.

The Treasury Board of Canada produces an annual report on the S&T performance of Canada, one which uses many of the standard OECD measures of R&D performance (Industry Canada 2008), but this report does not focus specifically on the performance of the health research system internationally. This report (the 'S&T Data Book'), covers federal spending on S&T, industry spending on S&T, higher education spending and research degree output, other human resources outputs, bibliometric outputs and some commercialisation outputs.

The main source of information for health R&D comparisons in Canada though is Statistics Canada, which collects information by surveying the various actors in health research across Canada (federal, provincial, higher education, business and private not for profit). Since all of these surveys are developed using the Frascati Manual (Organisation for Economic Co-Operation and Development 2002) they are all comparable to data collected in other OECD countries so can be compared easily at the international level. This is the role of the OECD itself, whose comparisons are so often used in analysing

the performance of Canada's R&D system, and something that we have already touched upon in the section outlining the position of Canadian health R&D in chapter 2.

CIHR evaluation program

As the main federal funder of health research in Canada, CIHR has a requirement for accountability and transparency. From its inception in 2000 (when it took over from the Medical Research Council of Canada), CIHR has always had some sort of focus on understanding its outputs. In its very earliest stages this consisted of tabulating the key results expected from CIHR research in three areas: Discovery and Knowledge creation; translation and transfer of knowledge to maximize the benefits of health research; and leadership and collaboration within the Canadian health research community (Canadian Institutes for Health Research 2001).

Pressure for increased accountability of federal R&D funds began in 2003, and by 2004 it was suggested to the House of Commons that the National Research Councils should strengthen their accountability for the outcomes of their research funding (Auditor General of Canada 2004). In their 2004 blueprint (Canadian Institutes for Health Research 2004), CIHR suggested that they would commit to evaluating the funding and work of CIHR. This blueprint set in motion a charter that would seek information on measuring ROI in health research from the Canadian health research community. By 2005 CIHR had acknowledged that to truly understand their outputs and outcomes, they needed a more formal evaluation strategy. As a result, CIHR convened two meetings in 2005 to develop a framework for measuring the impact of their funded research. These meetings brought together a group of international experts on measuring the impacts arising from health research, CIHR personnel and research funding organizations from around the world (Canadian Institutes for Health Research 2005).

The findings of these meetings, which covered a number of evaluation strategies used around the world for understanding the impacts of health research, were that the most appropriate approach for CIHR to use was a multi-dimensional approach that would capture impacts at many stages of research and across all four pillars of research. They decided to use the Payback Model described in chapter 3, as this allowed the evaluation of CIHR impacts to feed into accountability for its federal funds, as well as provided learning for how to better administer research support to enable maximum impact from the research (Canadian Institutes for Health Research 2005).

CIHR's version of the Payback Model uses the same framework as Buxton and Hanney's original version (see Figure 28), but this is combined with the CIHR institutes common logic model (Canadian Institutes for Health Research 2005) for modelling the impacts of funding provided to the 13 institutes, which models all research from CIHR goals through inputs, activities, outputs, methods for enabling outputs and outcomes through to the final impacts of research. An aspect of the payback model that has been modified by CIHR was the five impact categories. Originally, the five impact categories of the payback model[106] were rearranged to incorporate into the 'Economic Benefits' category two impacts normally categorised elsewhere: products developed in a commercial setting (formerly in 'informing policy and product development'); and cost-effectiveness of the health system (formerly in 'health and health sector benefits')(Canadian Institutes for Health Research 2005). More recently, the five categories identified by CIHR in 2005 have been modified to four categories where the 'research targeting and capacity building' category has been subsumed by the 'advancing knowledge' category (Figure 35).

---

[106] The payback categories adapted were those in the 2005 evaluation of the UK Arthritis Research Campaign (ARC) by HERG and RAND Europe {{135 Wooding,S. 2005}}.

Original Payback Category (Buxton, Hanney 1996)	CIHR 2005 Categorisation (Canadian Institutes for Health Research 2005)	CIHR 2008 Categorisation (CIHR 2008)
Knowledge Production	Knowledge Production	Advancing Knowledge
Research Targeting, Capacity and Absorption	Research Targeting and Research Capacity	Research Capacity
Informing Policies and Product Development	Informing Policy	Informing Decision Making
Health and Health Sector Benefits	Health and Health Sector Benefits	Health Benefits
Broader Economic Benefits	Economic Benefits	Economic Benefits

Figure 35. Development of CIHR payback categories

At the outcomes and impacts level of the common logic model, CIHR has identified expected outcomes/impacts from research that can be mapped onto the impact categories of the payback model. This is an important aspect of the CIHR method, the ability for their payback model to fit with their existing logic model. By successfully transposing the current logic model onto the new Payback model's impact categories, the transition in data collection and expected outcomes was managed with minimal disruption. Figure 36 shows how the two models fitted together.

CIHR common evaluation logic model   CIHR health research impact framework
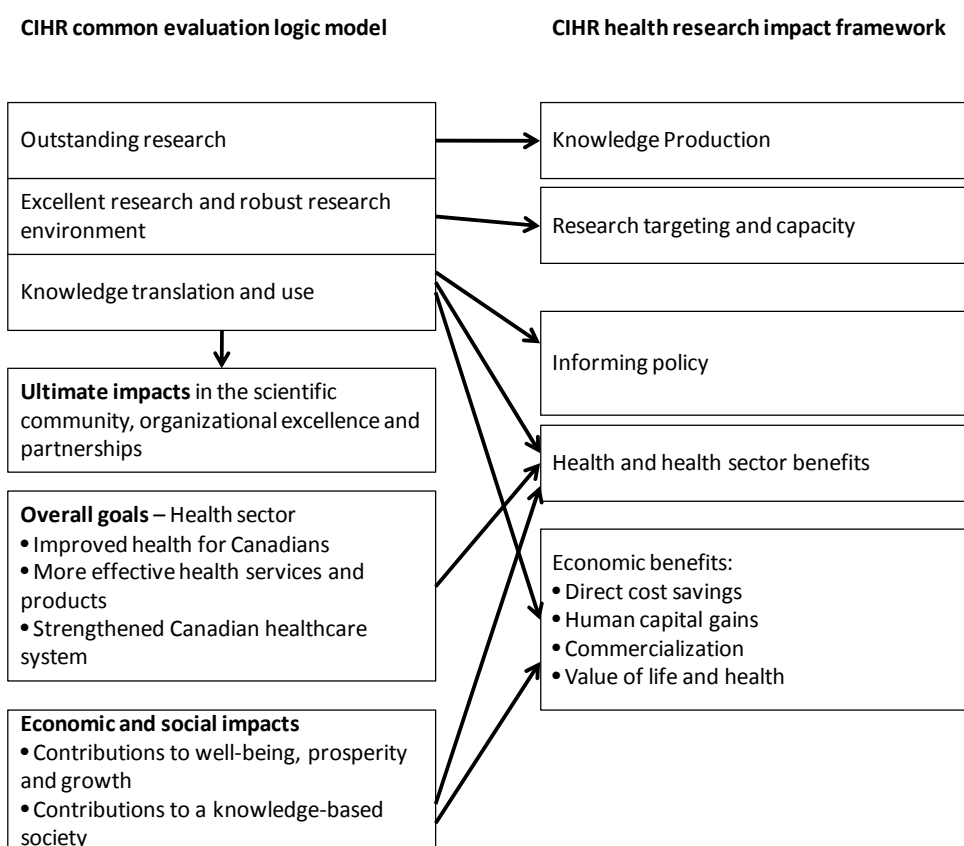
Figure 36. Relationship between CIHR's Common Performance Measurement and Evaluation Framework and the Health Research Impact Framework (As in 2005. With modifications to the logic model and payback categories

since then, specific wording may have changed for a 2008 version of this figure. However, this aptly illustrates the cross-overs between organizational logic models and payback categories using a real world example).

Source: (Canadian Institutes for Health Research 2005)

The impact categories on the right hand side of Figure 36 represent the areas in which CIHR have developed their own set of performance indicators, collecting data to monitor the impacts of the research they fund, Table 2 below shows the current set of indicators in each impact category.

Table 2. Performance indicators used by CIHR within their Payback categories.

Source: Recreated from presentation by CIHR to a Swedish Research Council meeting on evaluating return on investment in medical research, slides 8-11. (Canadian Institutes for Health Research 2007)

IMPACT CATEGORY	INDICATORS	DATA SOURCES
Advancing Knowledge	No. discoveries/ breakthroughs resulting from CIHR-supported research No. Canadian health research publications No. publications resulting from CIHR-supported research % Canada Research Chair (CRC) holders attracted to or retained in Canada No. and type of trainees supported by CIHR No. and type of Ph.D. graduates in Canada by year % Ph.D. graduates in Canada planning post-doctoral work in health	Bibliometric studies End of grant/research results reporting Program evaluations Databases of CRC holders Data available through Statistics Canada (i.e., census and survey data) Performance management data
Informing Decision Making	Impact of publications resulting from CIHR-supported research Impact of Canadian health research publications Research, policy and/or practice agendas influenced by funded research and/or CIHR institutes Clinical practice informed by CIHR-funded research Health system management decisions informed by CIHR-funded research Public policies informed by CIHR and CIHR-funded research	Citation impact analysis End of grant/research results reporting CIHR performance management data CIHR program evaluations Research user surveys Case studies (multi-method)
Health	Research study participants' health status directly affected by participating in CIHR-funded research Population health status influenced by CIHR-funded research Health-related quality of life influenced by CIHR-funded research Potential years of life lost (PYLL) for target disease categories (e.g. cancer, circulatory disease) influenced by CIHR-funded research	Case studies (multi-method) End of grant/research results reporting Statistics Canada data Studies to establish links to health research CIHR performance management data Analyses of publications
Economic	Number and nature of patents, spin-off companies and IP licenses influenced by CIHR-funded research Income from IP commercialization Commercial use of research funded by CIHR's commercialization programs Cost savings influenced by CIHR-funded research Human capital gains, including productivity influenced by CIHR funded research	End of grant/research results reporting Statistics Canada data Case studies (multi-method special studies) Technology assessment special studies Collaborative studies with Health Canada and Statistics Canada

There are a number of CIHR indicators in place that can be collected within their evaluation framework. It is not clear for all of these how to collect appropriate data, or indeed how to attribute it to the research findings (for example, cost savings influenced by CIHR research); however, since CIHR is one of the first funding bodies worldwide to have developed an ongoing health research evaluation system, this move towards standardized data collection within a framework for understanding impacts can be rightly lauded. CIHR is also in an enviable position of having funding designated for knowledge translation activities, something that has allowed it to develop and fine tune its evaluation strategy over the last 3 years.

Manitoba Health Research Council and Saskatchewan Health Research Foundation evaluation

At a provincial funding level, the expectations of funding impacts are different to those of national providers (be that at the level of research funder such as CIHR, or at the higher government levels of the Treasury Board). As such, any evaluation techniques used must reflect the questions most pertinent to provincial research funders. From discussions with different provincial agencies, it became clear during this project that there are some key aspects that provincial funders desire when evaluating their own funding.

Firstly, there is a focus on inputs to research and how this compares to other provinces. This links well with the national picture of research funding where the treasury department is keen to understand the standing of Canada in relation to other countries on input measures (Chapter 2). In the case of the Manitoba Health Research Council (MHRC), who we are using as an example of a provincial funding agency[107], they even funded an analysis of funding levels for health R&D across the provinces and how successful each was in 'pulling down' funding from CIHR (Birdsell, Asselbergs 2006). This analysis contrasted a number of input factors for provincial health research: funds spent on health research; funding mechanisms and major funding streams;[108] normalised CIHR funds in a province (by capita, comparison of per capita to national per capita spent, by percentage of total CIHR funds, and by pillar of research); and funds from the National Cancer Institute of Canada (NCIC). It is interesting to note that there are no genuine output factors for Manitoban health research in this analysis, although it could be argued that bringing funding into Manitoba from CIHR and NCIC could be considered an impact of research. This shows how difficult it can be to access output data in even the most accepted of forms; bibliometric data is available and able to be assigned at the provincial level, but it is an expensive process purchasing the data and analysis required to be able to use such data – often something that is not within the budget of smaller research funders.

In Saskatchewan, the provincial funding organization the Saskatchewan Health Research Foundation (SHRF) took a slightly different approach to understanding their research impacts. Rather than benchmarking themselves against the other provinces, SHRF contracted the Saskatchewan Institute of Public Policy (SIPP) to provide an economic impact assessment of provincial health research (Peach, Marshall 2008). This analysis used the four types of economic impact identified by the Payback Model (Box 5).

---

[107] The MHRC is funded through provincial government support grants and provides funding to health researchers across all four pillars within Manitoba. Their mission statement supports the concept that provincial funding groups have a slightly different remit than that of federal funders such as CIHR. The MHRC's role goes beyond research funding alone, with its mandate to also advise the Manitoba health minister in matters relating to health research.

[108] This includes the standard issues of whether a province funds people, projects, programmes and infrastructure, as well as specific initiative funding and knowledge translation.

Box 5. Saskatchewan's typology of economic impacts

Source: (Peach, Marshall 2008)

> - The direct and indirect economic effects of health research activity itself;
> - the effect of health research on productivity within the economy;
> - the impact of the commercialization of health research; and
> - the influence of health research on reducing health care costs or increasing health care effectiveness.

In the economic analysis, the direct and indirect research activity were calculated using input figures from Statistics Canada, an input/output analysis using the "North American Industry Classification System" (NAICS) multipliers for output in science and technology R&D, and a net impact by only using multipliers for output on those funds for health R&D coming from outside the province (since this is seen in this particular economic analysis to be equivalent to the use of in-province R&D funding for areas other than health, the other way to assess the net impacts of health R&D).

In terms of productivity benefits from research, Peach and Marshall (2008) assessed two main mechanisms for estimating the productivity impact. They addressed both the 'rate of return' approach and the 'value of human life' approach. Interestingly, they noted that both rely on a large economy and impacts of local research being felt locally, concluding that neither could satisfactorily show a value for Saskatchewan, where the benefits of health research are likely to be felt more widely outside rather than inside the province, implying that only a small proportion of benefit can be attributed to Saskatchewan.

With the commercial benefits of research, Saskatchewan has the problem of a small R&D base and a small industry presence making it almost impossible to assess commercial input to R&D. The report addressed this again by using a set of NAICS multipliers (although in this case multipliers for the return to the pharmaceutical and medical devices sectors rather than to R&D total) and this time using a hypothetical commercial input measure of $1m, to provide an example of the direct and indirect impacts for each $1m invested.

In the final category of reducing healthcare costs, the authors chose to only identify the healthcare costs within Saskatchewan, suggesting that increasing costs to healthcare provides opportunities for health research to improve efficiency. They did not attempt to identify any level of healthcare savings made through provincial research. There are additional 'economic benefits' identified in the report, although all of these are human capital measurements and represent jobs in health research, in the associated health and science sectors, and in terms of student numbers in health and other research courses. These numbers are not translated to any dollar value of economic benefit however.

Peach and Marshall's economic evaluation fits with the new evaluation framework put in place by SHRF in 2007 (Saskatchewan Health Research Foundation January 2007) in response to their own intention to evaluate their outcomes from the 2004 health research strategy mandated by the provincial Ministry of Health (Saskatchewan Health 2004). In the new evaluation framework SHRF build upon two well established evaluation techniques. Firstly, in order to ensure that their evaluations cover all the potential impacts health research can have, they have built upon the Payback Model. Then to ensure that the learning from evaluations is incorporated into the future actions of SHRF, they have included the implementation/evaluation model of Rogers and Freiberg (Rogers, Freiberg 1994) that in its simplest form takes an 'action and reflection' approach to implementation of evaluation findings (Figure 37).
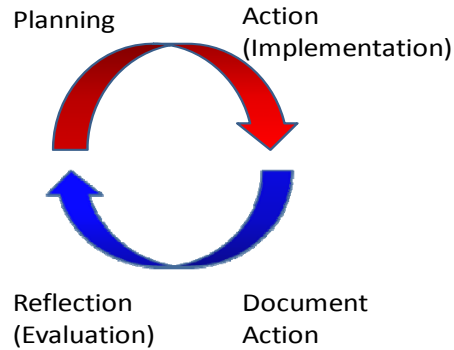
Figure 37. Action and reflection model incorporated into the evaluation framework of SHRF

Source: (Saskatchewan Health Research Foundation January 2007)

With the commissioning of the economic evaluation using the payback sub-categories of economic impact and the use of the payback model as a basis for the SHRF evaluation framework, we can see the importance of having an established framework in place for provincial funders. This provides an existing methodology to utilise, but also (in the case of the payback model particularly) allows the provincial funder to evaluate using similar criteria to important comparators, be that federal funders such as CIHR (also using the payback model) or other provinces (using Statistics Canada data on R&D funding and output measures).

National Alliance of Provincial Health Research Organizations (NAPHRO) evaluation

As shown by the MHRC example above, there is a definite desire amongst provincial research funders to be able to 'benchmark' themselves against other provinces and, ideally, against the country as a whole. To do this with research funding is relatively straightforward, using open-access figures that must be reported in the same way in order for Statistics Canada to use the data. However, if provinces want to be able to benchmark themselves against each other with respect to the outputs and outcomes of research, there is a need for an accepted model of evaluation between provinces.

NAPHRO has been in existence since 2003 as an umbrella organization meant to foster closer links between provincial health research organizations by promoting increased dialogue, linkages and partnership activities. One of the many areas that member organizations were interested in partnering on was research evaluation, and this area is now starting to come to fruition. NAPHRO have begun to address this issue by taking on board the work ongoing at the Fonds de la recherché en santé de Québec (FRSQ), who have been working on an evaluation framework that uses an impacts chain (a form of logic model) to understand the different stages of impacts that occur from health research (Figure 38).
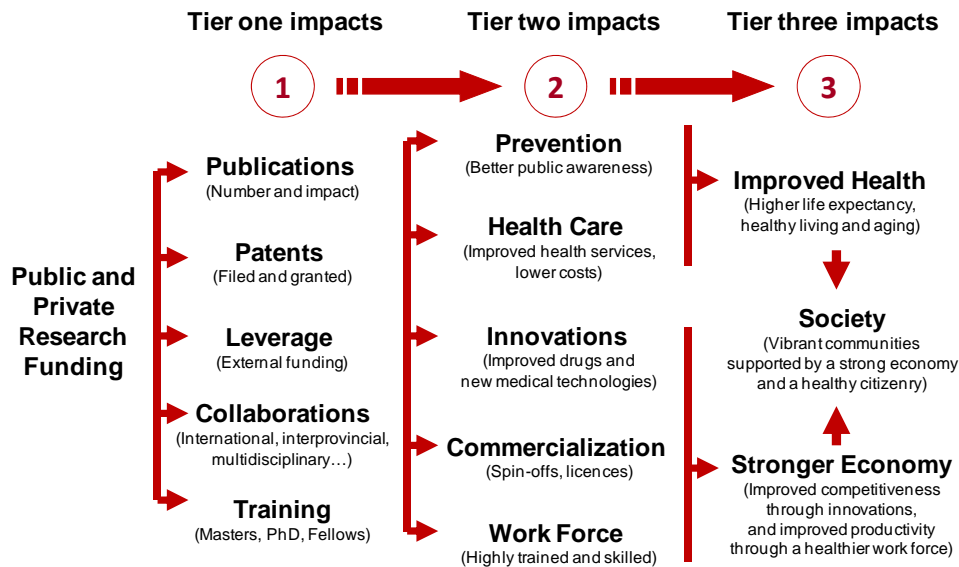
Figure 38. NAPHRO and FRSQ impacts chain model of research impacts (Beaudet 2007)

Reproduced with kind permission from the authors

Using this model, NAPHRO member organizations will be able to collect information into standardised impact groupings. Currently, the tier one impacts are the easiest to identify for funders, and some of the tier two impacts can also be readily identified (innovations, commercialisation and work force numbers). However, identifying data on prevention and healthcare improvements, and those impacts in tier three are a much harder process.

It is not expected that in the near future NAPHRO will implement the whole framework above, since there are many difficulties in understanding how research influences tier two and tier three (the attribution issue) and therefore, there are no accepted measures that could provide reliable data for the tier two and tier three impacts. NAPHRO are to use this impacts chain to aid their evaluation forum's work on understanding the publication and leverage impacts of the different provinces (Beaudet 2007), as has been done by the MHRC for example. There are already data sources in place for many of the publication and collaboration indicators, with the Canadian Bibliometric Database (CBD)™ forming the basis for bibliometric analysis. Constructed by the *Observatoire des Science et des Technologies* (OST) at the University of Quebec at Montreal, using Thomson Scientific databases (Science Citation Index (SCI), Social Sciences Citation Index (SSCI) and the Arts and Humanities Citation Index (AHCI)), the CBD lists publications by Canadian researchers according to discipline, institution, collaborative partners, etc. This database also supplies indicators on Canadian publications like volume of publications, international or sectoral collaboration, impact factor, and specialization index.

University Health Network (UHN) balanced scorecard

The UHN in Toronto has been using the Balanced Scorecard (BSC) to describe organizational priorities and measure progress since 2006. Their version of the BSC uses five modified domains: 'We' domain; Caring domain; Creative domain; Accountable domain; and Academic domain (Figure 39). Because the UHN is a network of hospitals, R&D performance is only one aspect of their role, and is only really measured through one domain of their BSC – the Creative domain; although the Academic domain includes education so could relate to research (University Health Network 2008a).

Figure 39. UHN BSC (adapted from: University Health Network n.d.**)**

In the 'Creative' domain, the UHN collects information on the numbers of citations, new clinical trials approved, number of disclosures, and the research budget variance (University Health Network 2008b). There is also a new additional measure being developed for turnaround times in the research ethics board process. Although this does a job of tracking certain impacts from research, it should be noted that it does not identify how these impacts arise, and is not a comprehensive impact assessment of the research conducted at UHN, merely an assessment of the outputs and processes of research.

Canada's Research-Based Pharmaceutical Companies (Rx&D) framework and evaluation

Although Rx&D is not a private sector funder of R&D, or indeed a producer, as an overarching body for the R&D interests of the private sector it is well placed to understand the impact of private sector R&D across all of its members, rather than within a single firm which would be the case for any evaluation process in a company. Rx&D's remit is wide; it is a grouping of all the research-based pharmaceutical companies in Canada with a mission to help create an environment within Canada that helps Rx&D better serve patient needs by discovering, developing, and delivering medicine. Rx&D is active on any aspect of the environment in order to ensure a highly productive nation in which to work and accomplish that mission.

With such a wide scope, it is perhaps unsurprising that Rx&D does not have a single approach to evaluating the impact of the pharmaceutical companies it represents; rather it has identified that there are sectors to which pharmaceutical R&D in Canada should be contributing. These areas are: improving health and saving lives; contributing to economic growth; improving the healthcare system; achieving increased employment; and investment in R&D for the future (Rx&D 2006b). Rx&D shows the role that Canadian Pharmaceuticals play in these areas, providing annual overviews of the current progress of the industry at a macro level, with selected examples of successful impacts from specific projects. These are presented as a 'factsheet' on the pharmaceutical industry in Canada and they identify the spending by Canadian Pharma; the value added by pharmaceutical employment (presented as an 'injection into the Canadian economy' (Rx&D 2007)); as well as other facts about the pharmaceutical industry in Canada and public perceptions of it. Rx&D has also released another factsheet that looks at the impact that pharmaceutical R&D has on health, entitled 'Canada's research based pharmaceutical companies: improving and saving lives' (Rx&D 2006a) which identifies the impacts that pharmaceutical R&D has on health in Canada. This identifies the changes in mortality and hospitalization in certain conditions, and the cost of drugs to the health system and attempts to link the two (albeit without actually identifying the percentage of any health change that is attributable to the drugs themselves). The factsheet also

cites a study carried out by an academic at Columbia University that shows that investment in new drugs reduces costs to the health system elsewhere (mainly through reduced disease burden).

It is interesting to note that the information that Rx&D produces about returns on the financial investment in industry does not follow the generally accepted standard evaluation procedures of companies themselves. It is no surprise to find that industry uses the balanced scorecard more often than government or non-profit funders, with its focus on customer and development processes (Balanced Scorecard Institute 2007). What is also apparent is that companies often have performance management systems for R&D that focus on the financial implications of industry R&D. For example, the costs of performance, timeliness of production, quality of product, and management of researchers and the research process (Nelsen 2008). These kinds of metrics and evaluation systems fit well with the overall performance frameworks we have described, but plant industry evaluation firmly in the camp of providing a positive return on investment (foregoing the longer view that pharma research should influence health and healthcare).

Canadian Health Services Research Foundation (CHSRF) evaluation

On top of the federal, provincial and industry funding for research, there also exist specific foundations across Canada whose role it is provide a focused funding stream in a specific research area. In particular, one of these foundations has played a large role in the understanding of the impacts of health research (in Canada and beyond). CHSRF has a mission to 'support evidence-informed decision-making in the organization, management and delivery of health services through funding research, building capacity and transferring knowledge' (Canadian Health Services Research Foundation n.d.b). This mission means that the understanding of the transfer of knowledge from research is key to their role in Canadian health research.

CHSRF have been an important driver of research evaluation in Canada since their role in understanding how to effectively transfer health research into organizational practice has allowed them to understand their own funded research and how best to implement their findings.[109] CHSRF use formal evaluations of their funding (project and program funding) as well as having formal evaluations of their organization. They have also built an evaluation framework complete with indicators for success of program funding, and claim to have built this into the actions of CHSRF (Canadian Health Services Research Foundation 2005).

The CHSRF evaluation framework builds on logic modelling techniques used for system evaluation (Canadian Health Services Research Foundation n.d.b) and allows for comparative collection of data across the different funding programs. CHSRF use their logic model as part of their international panel evaluation, conducted every five years by a group of international experts. This is the foundation level evaluation, although CHSRF also has program-level evaluations that are commissioned to evaluation experts. The two types of evaluation are dependent upon those performing the evaluation for their structure and focus rather than a standardised evaluation framework, even with the foundation level logic model used as a base for the international panel evaluation of the foundation as a whole. This becomes even more clear through the recommendations of the 2007 foundation level evaluation, which include 'routine external evaluation of its programs and more resources given internally to organizational and program evaluation' (Dussault, Davis et al. 2007).

In parallel, CHSRF also provide information for other organizations wanting to understand how best to use evidence in their own decision making – providing a framework for taking on board research

---

[109] For examples, see 'Insight and Action', issues 1 and 23 {{298 Canadian Health Services Research Foundation 2007; 299 Canadian Health Services Research Foundation 2007}}

findings rather than evaluating research findings (Canadian Health Services Research Foundation n.d.a). Although this is the complement to the research evaluation system, it provides valuable learning for the research evaluation system. The framework is a simple set of four headings that any organization wishing to utilise research needs to take into account (Box 6).

Box 6. CHSRF advice to organizations on using research in decision making (Canadian Health Services Research Foundation n.d.a)**.**

---

1. **Acquire evidence:** Can the organization find the research evidence it needs?

2. **Assess evidence:** Can the organization assess whether the research is reliable and high-quality, and whether it is relevant and applicable?

3. **Adapt its format:** Can the organization present the evidence to appropriate decision makers in a useful format, which synthesizes recommendations, conclusions and key issues?

4. **Apply it in decisions:** Does the organization have the skills, structures, processes and corporate culture to promote and use research evidence in decision-making?

It is important for us not to lose sight of this side of research impact, since this is one aspect of the impact that we are trying to identify in any research evaluation process. In a sense, this work provides a process map in successful research utilisation, one that can inform any attempt to understand the way research actually goes on to inform practice and policy.

Kidney foundation and Heart and Stroke foundation of Canada evaluation

There are many health research charities in Canada (and some could argue that the foundations above could be considered charities too). With over 3,800 registered charities with some link to health in Canada (including care organizations, hospices, and professional groups) (Canada Revenue Agency 2008), and with 25 currently involved in the Health Charities Coalition of Canada (Health Charities Coalition of Canada 2007), it is clear that health charities can have a profound effect on the health landscape. Although health research is only one of the ways in which these charities can affect the health landscape of Canada, it is not an inconsiderable one. Statistics Canada's most recent figures show that private-not for profit organizations account for half a billion dollars worth of health research funding in Canada, equivalent to about 8% of the total funding for health research in Canada in 2007 (Statistics Canada 2008).

In general charities do not have the resources with which to create ongoing systems for evaluating their impact, particularly since in most cases charity successes serve as an advocacy tool in attracting further funding rather than necessarily as a learning or accountability tool. The Kidney Foundation provide a good example of this, with research awards each year that highlight research excellence in Canadian (and world) kidney research. Although the process for selecting these 'excellent researchers' is not entirely clear (it involves peer assessment in some form), their impact upon the research community and the wider patient population is identified. As such, although this does not count as a formal evaluation, there are lessons to learn here from what can be considered as the impacts arising from charity funded research. For example, the most recent winner of the Kidney Foundation 'Medal for Research Excellence' (The Kidney Foundation n.d.) is shown to have produced a large number of publications; presented at many meetings; produced clinical guidelines; translated research into practice; and developing a new generation of clinical scientists. Even in an informal assessment of impacts, this provides useful information as to the potential impacts researchers can have, and also as to what research charities are interested in as outcomes of their funding.

The Heart and Stroke Foundation of Canada (HSFC) have evaluations on an ad hoc basis on specific projects that they fund, even getting researchers who can perform these evaluations as part of their research funding program. For example, HSF Ontario's Healthy Weights Area for Investment in Mission (HW AIM) Initiative is being evaluated as it progresses and HSF Ontario engaged the Alder Group to set up an evaluation framework that could monitor the progress of the initiative, evaluate the final outcomes and understand HSF Ontario's wider role in health promotion in Ontario. The request for proposals (RFP) for this particular funding showed the understanding present within HSFO about how to undertake evaluation of research. The stipulations in the RFP include a list of five deliverables shown in Box 7. These deliverables show that HSFO have a system in place to ensure that within any ad hoc evaluation, there is a structure to the evaluation process that can be easily understood and applied throughout the life of the evaluation.

Box 7. Deliverables for HSFO project on evaluating the HW AIM initiative (Heart and Stroke Foundation of Ontario 2007)**.**

> 1. Developing an overall **evaluation plan** that includes ethical considerations, a list of stakeholders, the key process and outcome questions that will be answered by the evaluation design, an outline of the evaluation methods, recommendations of the tools that will be used, and duration, phasing, and timing considerations. Evaluation plan will also include the setting of objectives and indicators for both process and outcome areas.
>
> 2. Developing a **logic model** that can be used as a management tool to identify progress of planned activities, and the relationships between planned activities and anticipated outcomes.
>
> 3. Developing an **evaluation framework** that matches each of the planned short-term and intermediate outcomes of the HW AIM with relevant indicators, methods of data collection, and data sources.
>
> 4. Develop, identify and/or adapt existing **data collection tools**, data analysis techniques, and a reporting framework.
>
> 5. Submitting a report summarizing above deliverables to HSFO by January 2008.

This is not the only example of the understanding of evaluation frameworks by HSFC; they are also involved in an international study of the impacts of cardiovascular research conducted across three countries. Project Retrosight (Heart and Stroke Foundation of Canada 2008) is using a payback framework, similar to that used by CIHR, to understand the impacts of particular case-study research projects from Canada, Australia and the UK from the last 15-20 years. This project will allow a greater understanding of the processes that lead to impacts from research as well as in international understanding of the importance of context in influencing research impacts.

Although these examples show that there is an understanding of how to effectively use evaluation and the ways in which these can be built around evaluation frameworks, they do not show any ongoing use of an evaluation framework at an organizational level. This is something that takes a resource commitment often beyond the means of research charities, whose funding accountability to donators means that, despite the obvious advantages of effective organizational funding evaluation (being accountable for the *outputs and outcomes* of the research to help achieve the charity mission), the charity often has a first requirement to be accountable for spending funding on scientific research, rather than administration (which is what evaluation is often seen as).

Practical Application: Examples of Frameworks in Use Internationally

Understanding the Canadian context is important in ensuring that any evaluation framework and metrics at the national level take into account the breadth of evaluation knowledge in the country, and also in ensuring that any developed metrics fit with the work currently underway in health research evaluation in Canada.

However, there is also a wealth of health research evaluation outside of Canada that can provide valuable learning for the country. Internal funders of research have been aware of this external expertise: CIHR used international expertise in the development of their evaluation framework, the HSFC are part of an international research evaluation project along with the UK and Australia, and industry funders are often international organizations who can pick and choose from the best evaluation systems for their industry in the countries they operate in. The following sections cover some of the current work ongoing and systems in place around the world; it does not cover all the work ongoing globally, but does give a flavour of the main areas of health research evaluation in other countries. The commissioned paper by RAND Europe in appendix A provides a more in-depth assessment of the evaluation frameworks and systems in place in organizations outside Canada than can be provided here.

Section One: United States

As the nearest neighbour, the USA is often the first comparison for Canada, and considering the USA's position as the country with the most health research citations, publications and the largest investment in health research in the world (King 2004), it is also the leader in health research.

In the USA, the evaluation of health research has taken a number of forms (unsurprising considering the size of the research investment). There is one stand-out study however that is routinely mentioned in any discussion on health research, and that is the 'Exceptional Returns' study supported by the Lasker Foundation and produced by Funding First in 2000 (Funding First 2000). Exceptional Returns was essentially an advocacy report sponsored by a health research organization that sought to show that the economic value that health research could have was wider than the traditional economic measures of sales, employment and cost savings, by monetarisation of improved health. The study was actually a selection of pieces of research supported by the foundation, but all used a 'willingness to pay' approach to valuing health – by identifying what individuals would pay for reduced risk of death; providing a value of around $3m to prevent death (statistical value of a life). This figure was then used to provide a monetary value for the drop in mortality seen in the USA from 1970 to 1990, once a proportion of the health improvement had been attributed to health research (in this case, using acute cardiovascular problems as a benchmark, they attributed 1/3 of all health gains to research). This indicated that research into cardiovascular disease had $500m return through lives saved, a return of around 20 times the initial investment. However, the Funding First approach has not been without its critics, with people claiming that the large value identified for the statistical value of a life, suggesting that regardless of the actual impact on health, the return to the research investment will always be sizeable, something that even the authors concede. The statistical value of a life attributed relies on a number of assumptions that can be questioned, including co-morbidity and the difficulty in assigning a dollar value to any health improvement in general. Also, the funding first approach makes the assumption that health gains in the USA are entirely due to research conducted in the USA. Perhaps for the USA this is not an unreasonable assumption, since such a large proportion of health research comes from the USA, but for any country attempting to reproduce this kind of study, there are significant issues to overcome with respect to the proportion of research from the country under study that can be reasonably have had an effect on health improvements (Nason, Janta et al. 2008).

In terms of ongoing research evaluation and evaluation frameworks it would be unwise to ignore the National Institutes of Health (NIH), who provide the majority of the US government's health research funding. In terms of evaluating impacts, there seems to be no NIH standard evaluation framework or process that is used across all funding. Instead, there are systems in place for different aspects of NIH's funding. For example, the funding for extramural programs is subject to prospective evaluation, with some centres being retrospectively evaluated too to understand whether they have achieved their goals (Manning, McGeary et al. 2004). The NIH also has a dedicated evaluation branch, part of the Office of Portfolio Analysis and Strategic Objectives. The role of the evaluation branch is to administer funding for evaluations, provide support to those performing evaluations and to publicise the results of the evaluations. The individual evaluations are qualitative studies that do not have a standardised framework, but do use the initial goals and aspirations of any study that were identified in the funding application. These are then collected and presented annually to Congress to show the outputs and outcomes of NIH research (Office of the Assistant Secretary for Planning and Evaluation 2007).

One overarching factor that is consistent within NIH is the need for peer-review of all funding applications. Even with such a large number of research applications put in each year, NIH insists on performing a peer-review process for all of them (although recently this has been streamlined to make it a manageable electronic process) (Scarpa 2007). This approach, combined with the details of evaluation above, show NIH's position on research evaluation in general. They consider the ability to make the 'correct' decision *before* funding as more important than understanding the actual outputs and outcomes of the research itself.

Outside of health research, the USA has embraced research evaluation, and a variety of different evaluation frameworks. This has been particularly evident in innovation circles, where there has been a lot of work on understanding the outputs and outcomes of funded innovation. Perhaps one of the best examples of this is the work for the Advanced Technology Program in 2003 (Ruegg, Feller 2003) which created a tool-kit for evaluation of R&D in the innovation field. This work focused specifically on the use of generic logic model frameworks that could be moulded to fit with the specific programs they were intended to evaluate. This work has been built on, with recent work on the evaluation of the US Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) which used logic modelling and diffusion theory[110] to evaluate program outcomes (Reed, Jordan 2007). In addition to logic modelling, the use of the balanced scorecard (Kaplan, Norton 1992) has also been suggested for R&D assessment, although these have tended to focus on the later aspects of R&D or innovation, where the financial and quantitative measurements more regularly associated with the balanced scorecard are easily identified ((Bremser, Barsky 2004, Jordan, Mote et al. 2006).

Section Two: United Kingdom

After the US, the UK is often the country that Canadian research is most compared with, and the work on health research evaluation has been steadily growing in the UK. This is illustrated by the presence of the 'UK Evaluation Forum' (UKEF), a collection of health research funders, conceived in 2003 and representing government, the commercial sector, private foundations and charities with the aim of improving evaluation knowledge for health research (Box 8).

---

[110] Diffusion theory is a broad systems theory that describes how technologies and practices are diffused and adopted, in the case of the Reed and Jordan analysis, to specific groupings of stakeholder.

Box 8. Strategic objectives of the UK Evaluation Forum (UK Evaluation Forum 2006)

> 1. To examine how member organizations carry out their own evaluations; to establish what information has been generated; and to consider how this has been presented.
>
> 2. Subsequently, to establish the lessons that can be shared within the UK, identify gaps in knowledge, research international evaluation practices and highlight challenges and opportunities for the future.

UKEF produced a 2006 report, 'Medical Research: Assessing the Benefits to Society', which pulled together different evaluation frameworks, methods and economic approaches to evaluating medical research. They investigated the use of the Payback Framework and the Royal Academy of Arts and Sciences in the Netherlands framework, which assesses societal impacts. Their conclusions suggest that there is no single best way to evaluate the impacts of health research, and that there are also no 'best indicators'; however they do identify what any indicator should capture (Box 9).

Box 9. UKEF description of the requirements for indicators of societal impacts (UK Evaluation Forum 2006)

> - Capture all relevant research activity or be representative of that activity.
>
> - Where appropriate, link outcomes and impacts to the original objectives of the funder.
>
> - Allow for the incremental and cumulative nature of research, alongside the timescale of scientific progress and its inherent uncertainties.
>
> - Demonstrate validity and reliability.
>
> - Provide an efficient means of capturing information to avoid using resources that might otherwise be devoted to new medical discoveries or their application.

In a recent follow up, the UKEF have commissioned a piece of research to understand the economic impact of UK health research from a consortium of research organizations (the Health Economics Research Group at Brunel University, RAND Europe and the Office of Health Economics). This has used a case study approach to this question, with Cardiovascular Disease (CVD) as a case study, since this is an area of research that has been well investigated in terms of understanding the links of research to health benefits. This work used a number of innovative techniques to try and create a bottom-up method for understanding the attribution of impacts to research that would be more descriptive than previous attempts at economic valuation of impacts. These included understanding fully the inputs to CVD research, analysing the way UK research is cited in clinical guidelines (as a method for estimating the influence of UK research on UK health), a QALY based approach to understanding the monetarisation of health impacts,[111] and an investigation of the spillover effects of from public research (Buxton, Hanney et al. 2008). The report also goes on to test the model created that takes into account all of these economic impacts, with a more difficult case study area (mental health research) where there is not as strong an understanding of the way impacts arise from research. This attempt is seen by the authors as moderately successful, although with the major caveat that the data that is so readily available for CVD is not there for mental health.

A parallel piece of research, funded by one member of UKEF (the Wellcome Trust), has used a more traditional econometric approach to estimating the return to the UK from investment in health research.

---

[111] This approach was driven by evidence on the effects and costs of specific research-derived interventions, not from macro-level, temporal changes in mortality or morbidity.

This work, by the London School of Economics (work in progress), builds on their 2007 research assessment that used the Murphy and Topel approach to estimating economic benefits with a willingness to pay approach (McGuire, Raikou 2007). In the most recent research, the use of the willingness to pay approach was again the centre of the analysis, and this top-down approach has provided an overview of the whole system and a monetary value for the entire health research system.

Organizations in the UK are also building evaluation frameworks into their reporting systems, something made slightly easier by the general requirement from funding organizations for researchers to report on the outputs of their funding through 'end of grant' reports. This is now being utilised by the Arthritis Research Campaign (ARC), the most forward looking of UK research charities with respect to impact evaluation.[112] ARC are now attempting to use the payback framework as a way to incorporate impact analysis into their organizational processes, working with RAND Europe and Brunel University to create a questionnaire that can be answered by grant holders at the end of their funding and again at a later point (such as five years after the end of the grant). This works well for a research organization such as ARC, where many researchers are repeatedly funded ensuring they have a reason to be involved in impact analysis so long after the end of a grant. The output of the questionnaire is aggregated data on the many outputs and outcomes from the grants that can be analysed along a large number of lines (such as applied versus basic research, projects versus fellowships, early versus late career stage researchers, etc.) that can then inform ARC as to how best fund research to achieve particular types of outcome (ARC reference).

The Medical Research Council (MRC) and Wellcome Trust have also instigated evaluation practices into their own organizational set-up. The MRC produce an annual report on the selected highlights of their published research (Medical Research Council 2008b), as well as ad hoc evaluation reports on programs of research (Medical Research Council 2008a). The MRC are also currently working with RAND Europe looking at ways that they can best evaluate the outcomes of their programs of research (Wooding 2008a). The Wellcome Trust are also active in evaluating the impacts of their programs, and have identified eight categories of impact that their research can have (Wellcome Trust)(Figure 40), provide evaluation advice for grant holders on what frameworks and methods are useful for self-evaluation of projects (Wellcome Trust), and they are developing indicators of research impact as part of their strategic plan (Wellcome Trust 2005).
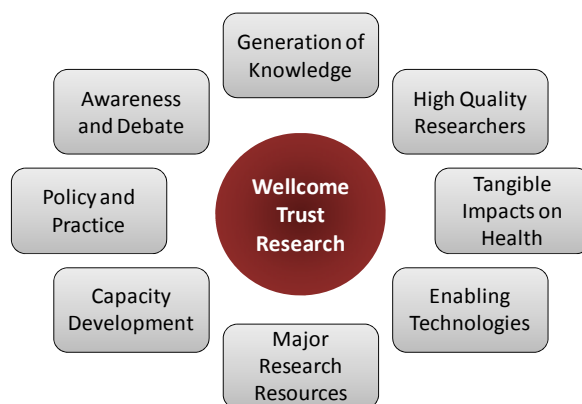
Figure 40. Wellcome Trust impact categories for research outcomes (adapted from text in: Wellcome Trust n.d.**)**

---

[112] ARC have been working to understand the impact their research can have since around 2003, with an in-depth study on the payback from case-study grants that provided a large amount of information on the different ways research can have impact {{192 Wooding,Steven 2004; 135 Wooding,S. 2005; 156 Hanney,S.R. 2004}}.

Section Three: The Netherlands

As mentioned in the UK section, the Netherlands are also taking the issue of impact analysis for health research seriously and are producing a number of interesting frameworks and evaluations that further the field. The Netherlands public organization for health research and development, ZonMw, have been involved in furthering the research evaluation field in Holland through their own evaluations of programs in the country and their willingness to address different evaluation frameworks in understanding the outcomes of their own work. For example, ZonMw supported a payback study to evaluate the Healthcare Efficiency Research programme (Oortwijn, Hanney et al. 2008) which created payback profiles for different projects and concluded that there were a number of difficulties in using the payback approach on grants that had ended recently and with independent scoring of research outcomes (used to create the payback profiles for projects).

ZonMw have also investigated the home-grown approach of societal impact evaluation of research (Council for Medical Sciences 2002). In this approach, there is a matrix of indicators that provide information on the societal impacts of the research – with a focus on how research is communicated to different stakeholder groups (Figure 41). Indicators that work within this matrix are actually very similar to indicators identified through other evaluation frameworks with publications, citations and funding forming the basis for evaluations.

Knowledge Impact	Products	Exchange and esteem	Use	Knowledge Societal Sector
Social		Indicators		Public Sector
Economic				Private Sector
Cultural				General Public
Attractiveness (revenues generated)				

Figure 41. Societal impact of research – matrix of indicators (adapted from van Ark 2007 with kind permission from the authors)

This work was originally put together for the Leiden University Medical Centre (LUMC) and continues to be evaluated there (see appendices A and E for full details on the LUMC approach). LUMC have been able to implement an IT system for monitoring the progress of each department in the medical centre, although in the 2006/7 roll out of the system, fewer than half the departments responded to the questionnaire, none of the departments fully understood the system (often reporting what they wanted to report rather than what should have been reported), and the majority of indicators focused on process rather than impacts (Ellenbroek 2007).

The societal impact approach has become a viable business opportunity in the Netherlands, and the research consultancy SciQuest has been providing advice and tailored evaluation processes in the Netherlands since 1995. The current SciQuest method utilises a Research Embedment and Performance Profile (REPP) which houses indicators in five domains: Collaboration and visibility; Science and Certified Knowledge; Education and training; Innovation and professionals; and Public Policy (SciQuest n.d.). Indicators in each domain are scored and then plotted onto the REPP to provide a visual representation of the output/outcomes of the group under evaluation (Figure 42).
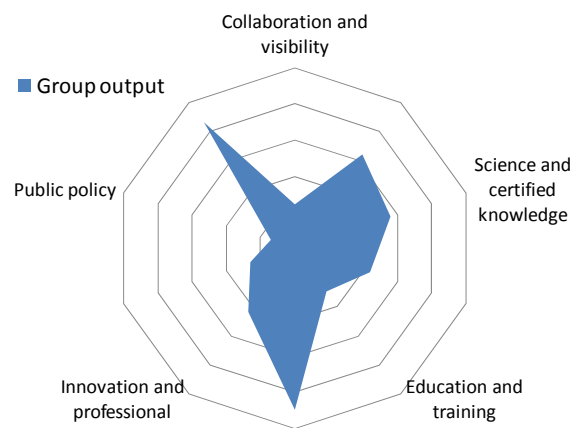
Figure 42. Example of a REPP from the SciQuest website (based on: SciQuest n.d.)

In the Netherlands there is an understanding of the need for evaluation, with work ongoing outside of the major national public sector funders of research. For example, in 2007 the association of Dutch Health Charities held a one day workshop on research evaluation and the various approaches that can be taken for charities (including bibliometric analyses, the societal impact method and the payback framework). The Netherlands also houses one of the world's foremost bibliometric analysis groups at the Centre for Science and Technology Studies in Leiden (CWTS); a research group involved in national and international evaluations of research impact through bibliometric analysis (see the CWTS website for further details of their ongoing work - www.cwts.leidenuniv.nl)

Section Four: Sweden

In parallel to this work for CAHS, the Swedish Research Council (SRC) are also investigating the ways to evaluate health research; in their case with a particular focus on the economic impacts of health research. This work was kicked-off with a meeting held in Sigtuna, Sweden that brought together stakeholders in health research evaluation from around the world, with a particular focus on Scandinavian countries funding organizations (Billig 2007). Through this meeting, the Swedish Research Council has convened a working group of international experts that is investigating the different methods and frameworks in place for understanding the economic impact of health research. This group is expected to produce a document by January 2009 that will dictate the direction that the SRC take their own future evaluation systems.

Prior to the 2007 meeting, the National Institute of Public Health (NIPH) set out in their 2003 strategic plan (National Institute of Public Health - Sweden 2003) their new database to map the work on public health in Sweden (funding, human resources, and descriptions of projects). This database was intended to be the basis of an international evaluation of Swedish public health research. This evaluation used an international expert panel to evaluate interview findings and some data from the database, but used no specific framework to build the evaluation (Kamper-Jorgenson, Arber et al. 2004). The SRC also investigated the 'medical research situation' in Sweden, investigating the impact of funding changes in Sweden on human resources in medical research and scientific publications (Vetenskapsrådet 2004) .

Bibliometrics is used in Sweden by both the SRC and the most famous of its universities, the Karolinska Institute, in understanding the impact of their research findings. For the SRC, 2007 saw the production of a bibliometric analysis of all Swedish scientific research from 1982-2004 (Karlsson, Wadskog 2007).

This aligns with regular reviews of the bibliometric outputs of Swedish health research (Tiessen 2007). The Karolinska Institute also produce their own evaluations using bibliometric analysis, and have developed their own bibliometric institute in 2006 to facilitate analysing the university's performance (Karolinska Institutet 2008).

Section Five: Australia

Australia has been one of the most active countries in recent years on assessing the impacts of research, particularly in terms of understanding metrics systems for evaluation. The Australian National Health and Medical Research Council (NHMRC) set up a performance measurement framework in 2003, finalised in 2004, that allows tracking of the outputs and impacts of grants through requiring researchers to conform to specific end of grant reporting categories (Figure 43). Within each outcome, lie a number of sub-groups and indicators that provide information for NHMRC and a template for evaluations and data collection (National Health and Medical Research Council 2004). This framework is used to organise the findings from the evaluation of end of grant reports on an annual basis (for example (National Health and Medical Research Council 2007)).
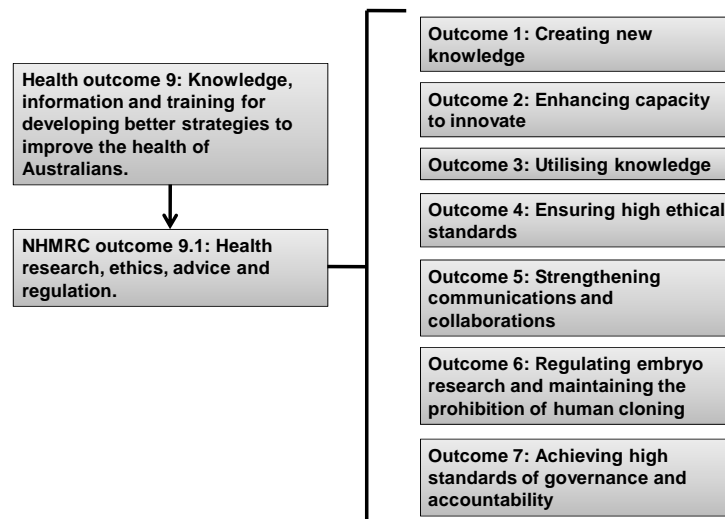
Figure 43. NHMRC outcome and output framework 2003-2006 (adapted from (National Health and Medical Research Council 2004)**)**

At the Australian National University (ANU) there is a designated program of research to investigate the impacts of research, the Research Evaluation and Policy Project (REPP), whose major focus is on "research on the advanced quantitative analysis of scientific performance and the organizational structure of Australia's research landscape" (Research Evaluation and Policy Project 2008). REPP has a role in producing indicators for the Australian Research Quality Framework (RQF), the way in which block funding for universities in Australia will be determined. They also produce a large number of reports that illuminate the different sorts of indicators that can be used to evaluate research (Research Evaluation and Policy Project 2005).

Australia also produced its own version of the Funding First report in 2003, when Access Economics conducted a willingness to pay economic benefits study on Australian health research (Access Economics Australia Economic Consulting 2003). This used essentially the same methodology as the US version by the Funding First (Funding First 2000) but modified the way in which Australian R&D affected Australian health, by approximating the impact of Australian R&D on Australian health to be the same percentage as Australia's share of the world's R&D (2.5%), by introducing a quality of life economic value using DALYs, and by increasing the percentage of health improvement attributable to research to

50% from 33% (Access Economics Australia Economic Consulting 2003). Even with these changes, the returns found showed a strong return, with $5 return for every $1 spent.

Section Six: Other Countries

Although we have highlighted a number of countries here, the work on evaluating the impacts of research is happening across the globe. Many countries have taken to using the Payback Framework in one form or another: these include research systems as diverse as Hong Kong and the Republic of Ireland (Nason, Janta et al. 2008); other countries and regions are bringing on their own evaluation methods with Germany (Gerhardus, Dintsios 2005) and Cataluña (Berra, Guillamón et al. 2006) both utilizing versions of evaluation frameworks to assess their activities. New Zealand's Health Research Council investigated their impact in 2004 using an international benchmarking of R&D inputs, but with additional background on the research systems in place in the benchmark countries (Garrett-Jones, Turpin et al. 2004). In Africa, the Afro-Nets networks for health research and development have evaluation as a key issue to address for the continent and producing documents on self-evaluation (LaFond, Kleinau et al. 2001); whilst in Latin America, there is also an acknowledged need to understand the outcomes of health research (Council on Health Research for Development (COHRED) 2006).

It is not possible to showcase every country and evaluation that is ongoing in this document. However, the selected countries give a flavour for those places where the most work is taking place in health research evaluation. These selections should provide a sufficient background for the reader to understand what the main issues within evaluation are and highlight the importance of understanding the impacts of research for all funders around the world.

Learning from Experience: Developing a Framework for Canada

Tying together the two sections above, we can see that there is a wealth of evaluation material in Canada and the world beyond. Some aspects have been more universally accepted than others. In terms of evaluation frameworks, the Payback Framework is something that is used all around the world and is now being built into organizational reporting systems, suggesting that there is a high level of acceptance of the way it groups and categorizes impacts as well as providing a framework within which to collect those impacts.

It must be stressed however, *that the most important aspect of any evaluation is not the framework itself but the data that can be collected and analyzed within the framework*. Therefore, the Payback Model provides a method for collecting data and there is no reason why this data could not be identified and collected using another framework (such as the societal impacts framework in the Netherlands, or the balanced scorecard). Once the data exists (and there are important caveats about the data collected that have been discussed in chapter 3), it can be easily compared with the same data collected through a different framework. Perhaps the best example of this is citation data, which is present in all frameworks of research impact since it is easy to collect and well understood. Citation data collected within a framework based on the Payback Model is the same as citation data collected within a Societal Impacts framework, and therefore entirely comparable. The important aspect of evaluation is having the data to produce a good evaluation, preferably one that can then be comparable to evaluations of similar organizations or projects to allow the most learning to arise from the evaluation.

There are some aspects of evaluations, frameworks and indicators that are agreed upon across the world. Despite occasional questions regarding its validity, bibliometrics are now a generally accepted measure of research impact in the academic sphere; so much so that the English Department of Health even now uses it to aid major funding decisions (Van Leeuwen, Grant 2007) and the Australian government is using it to distribute funds to universities through the RFQ. Collecting information on

funding, patent data, and research capacity (in terms of researchers and research degrees) are also all fairly standard indicators across all systems.

The further from the research itself an impact occurs, the more difficult it becomes to find agreed indicators, so the fact that the metrics and data collected in different evaluations for impacts on health, policy and the economy are the most discussed and argued comes as no surprise. Evaluators tend to approach this problem in one of two ways: either they decide to only evaluate the immediate outputs of the research using well established indicators of knowledge production, or they make assumptions that underpin indicators for further reaching impacts.

Many of the international comparisons of research use the first approach, since it is simple to access information through the OECD and other international data repositories. At the national and organizational level, those evaluations that attempt a further reaching impact assessment make use of assumptions to provide information for a fuller evaluation.

Not least amongst these assumptions is the attribution of further impacts to research (see Appendix D). The attribution issue is one to have dogged many projects, and has been a major criticism of the 'Exceptional Returns' studies undertaken in the USA and Australia (who despite using the same methodology use different percentages for the attribution of health impacts to R&D, based on a combination of assumptions. Other assumptions that regularly appear and are consistently questioned include those used in valuing health gains (the assumptions that underpin the $3m value of a life are regularly discussed), the assumed time lags between research and impacts, and the ways in which impacts sum and the avoidance of double-counting impacts.

One area that is often missing from research evaluations is that of funder balance. Considering most evaluations are conducted by research funders themselves, it is no surprise that there is not a balance of different research funders in evaluations. However, attempts by countries to evaluate their R&D efforts regularly focus on one type of research funder, to the detriment of the others (most commonly focusing on the wider impacts of public funding, the sales or turnover of the private sector and often ignoring the private not-for-profit sector entirely). The most recent work in the UK for the UKEF on the economic impact of CVD and mental health research has tried to address this issue however and investigated the full inputs to research and how they interact to produce impacts (Buxton, Hanney et al. 2008).

By understanding what has worked well and what is commonly questioned we can add value to any system and metrics for Canada in two major ways:

1. We can ensure that we use the best existing metrics and not fall into the traps that other evaluations have found criticism in.

2. We can identify where a new framework and or metrics can add most value to the science of 'research on research' that is becoming more and more important globally.

Evaluation Methods: Collecting Data

All of the above frameworks provide a method for understanding what should and should not be collected for an evaluation of health R&D. What they don't necessarily do is dictate how that data should be collected. For any attempt at comprehensive evaluation, there must be a number of different methods used in order to triangulate evaluation findings and cover the full range of impacts that health R&D can have (Ruegg, Feller 2003). Evaluation methods can be broadly broken down into two groups, quantitative and qualitative methods.

Quantitative methods

In order to understand the impacts of research and be able to simply compare one type of research to another, the easiest approach would be to provide a numerical impact score for each evaluation. Although this is not possible for all impacts, some evaluation methods do allow this sort of easily comparable counting of impacts. Quantitative methods, because they concentrate on data that is numerical provide data that can be statistically analysed, but are also necessarily narrow in their focus. In evaluating health R&D there are a number of quantitative methods involved, each of which has benefits and drawbacks.

The most widespread quantitative measure is **bibliometric analysis**, the study of publications and citations. Bibliometric analysis works on peer-reviewed publications and can provide information on the output of researchers (counts of publications), the connectivity of researchers (co-publication analysis), and the scientific impact or quality of publications (citation analysis). The quality measure for citations is based on the premise that scientific citation reflects the peer-esteem for the publication, based on its scientific merit (UK Evaluation Forum 2006, UNESCO Institute for Statistics 2005). These measures all reflect the knowledge impact of the research or researcher, but more recently bibliometric analysis has been used to address the impact on decision making of clinicians by citation analysis of clinical guidelines and medical education (Buxton, Hanney et al. 2008, Grant, Cottrell et al. 2000)(Webster, Lewison et al. 2003), the impact on the general public by citation analysis of newspaper and media articles (Webster, Lewison et al. 2003), and analysis of researcher CVs to track researcher mobility (Sandström 2008) . Bibliometric analysis has critiques (covered well in (Moed 2005)), which fall into four main groupings: comparisons between disciplines may be difficult since disciplines have different publication and citing behaviour; attributing a publication to a particular piece of research funding is difficult; getting bibliometric information and analysing it is expensive; and a focus on research publications can lead to neglecting other methods of dissemination.

Related to bibliometric analysis is **technometrics**, the study of patent citations. This provides an opportunity to see how research informs innovation, and also how patents inform other patents through analysing the publications cited in patent applications (Moed 2005).

Using **surveys** of researchers, clinicians, policy makers, etc. can provide quantitative information on research outputs from the point of view of different stakeholders in the process of research translation. Depending on the survey, this can also provide qualitative information, if open questions are used. Surveys can access information that data-gathering cannot since they go to the actors in the research translation process itself, but they can introduce an element of survey bias, missing unexpected impacts. If designed to collect quantitative data then they may also miss contextual information that could be important in understanding the data. The more detailed a survey is, the more resource intensive it becomes, but having detailed quantitative surveys can be very useful in understanding the wide range of impacts arising from research. Currently, the Arthritis Research Council (arc) in the UK are producing a detailed survey to provide quantitative information on the impacts of all their individual research funding awards (Wooding 2008b).

Economic rate of return analyses have been popular with advocacy groups, policy makers and governments since they provide a dollar value to the impacts that research has. These use a variety of tools to understand the monetary impact of research findings, and generally align along four types of impact (Buxton, Hanney et al. 2004):

- Direct cost savings to health care systems.

  Either through improved efficiency of practice, cheaper treatments (rare), or by reducing the burden on the health system (through preventative measures). These analyses can suffer from difficulty attributing impacts to research and accounting for the cost savings because of one aspect of research (e.g. is it immunisation or public health awareness programs that reduce influenza outbreaks?)(UK Evaluation Forum 2006). Also, HTA is away to understand the improved efficiency due to new research already commonly in use in Canada (Canadian Agency for Drugs and Technologies in Health 2008).

- Benefits to the economy from a healthy work force.

  Improved health means that individuals are able to work more often and for longer. This provides clear GDP effects through increased production. This 'lost production' approach uses a human capital method that identifies what is lost from the economy through illness, but does not take into account replacing that productivity with unemployed people (UK Evaluation Forum 2006).

- Benefits to the economy from commercial development.

  When considering the return on investment in monetary terms, commercialization is the most obvious route to economic impact. Commercialization can be measured in terms of industry profits, industry collaboration with academics, the creation of spin out companies and inward investment, as well as proxy impact measures such as patenting and licensing (UK Evaluation Forum 2006).

- Broader benefits to society of the health gain from medical advances.

  Aside from improving productivity, better health in society can be seen as an intrinsic benefit to society. If it is possible to place a monetary value on improved health, then the true value of health research could be assessed. This is the route that the 'exceptional returns' studies (Funding First 2000, Access Economics Australia Economic Consulting 2003, Rosenberg 2002, Murphy, Topel 1999) have taken in estimating the monetary return on health investments. By defining a value for elongated life or improved health through a willingness to pay methodology (where people are asked what they would be willing to pay for discrete improvements in health), it is possible to determine the dollar value of improved health. To assign this to research, studies assume a certain percentage of health increases are due to research, allowing for a value of health improvement due to research. Examples of this method are discussed in the following chapter.

Qualitative methods

Complementing the quantitative measures that provide easily comparable data on research impacts are a number of qualitative research methods that provide rich contextual data that allows a greater understanding of how research impacts occur and is able to identify unexpected impacts as well as the more common outputs of research.

To gain the fullest picture of a piece of research, it is necessary to gather as much information as possible on the research project in question. The most common way to do this is through **case study**

analysis. Using case studies enables attribution to be determined with more certainty than data analysis, since they can follow research to their effects. However, since case studies are of individual research projects they are not easily generalizable and may not be reflective of research as a whole. Within case studies, a wide variety of methods might be used to gather information, such as interviews, data and literature reviews, and bibliometric analysis (for an example see Wooding, Anton et al. 2004). A slightly modified version of the research case study, is that used by the Wellcome Trust in the UK and the National Institute of Health and National Science Foundation in the USA, who use narrative case studies that focus on a particular breakthrough or innovation and then identify the factors that allowed that to happen (Ruegg, Feller 2003, UK Evaluation Forum 2006).

Commonly used to assess research applications prior to funding, and the key aspect to quality control in journal publications, **peer-review** is a well respected and commonly used method for addressing the quality of a scientific submission, but is not regularly used in terms of impact assessment. In the UK, it is common practice for those receiving research funding to file end of grant reports that are reviewed by the funding agency. However, since these are required at the end of the funding period, often not enough time has passed for impacts to occur from the research (for example, citations require a 3-5 year window for assessment). In the US, peer-review is still considered to be the main way to evaluate the merit of science (Scarpa 2007, National Academy of Sciences Committee on Science, Engineering and Public Policy 1999).

Figure 44. Available methods for evaluating health research impacts (adapted from (UK Evaluation Forum 2006)

Method	Pros	Cons
Bibliometric analysis	• Quantitative: measures volume of output • Can be used to indicate quality of output • Enables analysis of global trends • Suited to repeated analyses • Can be applied to patents (technometrics) • Being developed for use with impacts as well as outputs	• Data are difficult to compare across research fields and disciplines • Analysis complicated by the introduction of electronic publications and open and public access journals • Expensive to collect data and analyse • Only able to investigate peer-review academic publications
Surveys	• Can identify outputs and outcomes associated with particular pieces of funding/research • Can provide qualitative analysis of outcomes, e.g. quality of trained researchers, business/academic interactions.	• Dependent on contact details being available, e.g. for past award holders • Poor response rates can lead to biased responses
Economic rate of return analysis	• Can be applied to variety of sectors • Can be used comparatively, e.g. contribution of cost effectiveness studies • Quantitative • Provides big picture and context • Potentially powerful political tool	• Involves subjective decisions of what's involved and therefore what to 'cost' • Difficult to put financial value on many influences involved • Heavily depend on monetary valuation of non-monetary goods (e.g. quality of life) • Difficulty to identify contribution of individual funder/sector/country
Case study analysis	• Provides in-depth analysis of the process of discovery • Can demonstrate pathways from research to application and impact • Information useful for a range of purposes (e.g. reporting to stakeholders, media) resource intensive	• Selection bias: cases chosen may not be representative • Often difficult to track and interpret the history of scientific discovery • Problems of recall bias • Method can be highly resource intensive
Peer review	• Well understood component of research management • Widely accepted by the research community	• Time consuming for experts • Concerns about objectivity and variability of judgements and lack of transparency

References

ACCESS ECONOMICS AUSTRALIA ECONOMIC CONSULTING, 2003. Exceptional Returns: the value of investing in health R&D in Australia.

ASSOCIATION OF UNIVERSITIES AND COLLEGES OF CANADA, 2006. *Measuring Success in Relation to Federal Investments in University Research: An AUCC Discussion Paper.* Ottawa, Ontario: AUCC.

AUDITOR GENERAL OF CANADA, 2004. *Report of the Auditor General of Canada to the House of Commons: A message from the Auditor General of Canada - main points.* Ottawa, Ontario: Office of the Auditor General of Canada.

BALANCED SCORECARD INSTITUTE, 2007-last update, balanced scorecard examples & success stories [Homepage of Strategy Management Group], [Online]. Available: http://www.balancedscorecard.org/BSCResources/ExamplesSuccessStories/tabid/57/Default.aspx [July 31st, 2008].

BEAUDET, A., 2007. Return on Investments in Health Research: Defining the Best Metrics - A Provincial Perspective. Presentation to CAHS Third Annual Meeting edn. Montreal, Quebec: .

BERRA, S., GUILLAMÓN, I., MÉNDEZ-VÁSQUEZ, R., SUÑEN-PIÑOL, E., SÁNCHEZ, E. and PONS, J.M.V., 2006. *Impact assessment of the TV3 Telethon on biomedical research in Catalonia.* Barcelona, Catalunya: Fundació La Marató de TV3: Agència d'Avaluació de Tecnologia i Recerca Mèdiques (AATRM).

BILLIG, H., 2007. Workshop on "Economic Returns of Medical Research". Presentation to the Swedish Research Council Workshop on Economic Returns of Medical Research edn. http://www.vr.se/download/18.5ec4ab911161ebf14dc80005414/H+Billig+Introduction.pdf: .

BIRDSELL, J.M. and ASSELBERGS, M., 2006. *Health Research Funding In Manitoba: A Description and Comparison Across Canada.* Manitoba, Canada: On Management Health Group.

BREMSER, W.G. and BARSKY, N.P., 2004. Utilizing the balanced scorecard for R&D performance measurement. *R&D Management,* **34**(3), pp. 229-238.

BRIMACOMBE, G.G. and GRUENWOLDT, E.C., November 2007. *Moving at the Speed of Discovery: From Bench to Bedside to Business.* Ottawa, Ontario: Association of Canadian Academic Healthcare Organizations.

BUXTON, M., HANNEY, S. and JONES, T., 2004. Estimating the economic value to societies of the impact of health research: a critical review. *Bull World Health Organ,* **82**(10), pp. 733-739.

BUXTON, M., HANNEY, S., MORRIS, S., SUNDMACHER, L., MESTRE-FERRANDIZ, J., GARAU, M., SUSSEX, J., GRANT, J., WOODING, S., NASON, E., ISMAIL, S. and KAPUR, S., 2008. Medical Research – What's it worth? Estimating the economic benefits from medical research in the UK. Report to the UK Evaluation Forum edn. London, UK: .

BUXTON, M., HANNEY, S., PACKWOOD, T., ROBERTS, S. and YOULL, P., 2000. Assessing Benefits from Department of Health and National Health Service Research & Development. *Public Money and Management,* **20**(4), pp. 29-34.

BUXTON, M. and SCHNEIDER, W., 1999. *Assessing the Payback from AHFMR-funded research.* Alberta, Canada: Alberta Heritage Foundation for Medical Research.

BUXTON, M.J. and HANNEY, S.R., 1996. How can payback from health services research be assessed? *Journal of Health Services Research Policy,* **1**(1), pp. 35-43.

BUXTON, M.J. and HANNEY, S.R., 1994. *Assessing payback from Department of Health research and development: second report. Volume 1: the main report.* HERG Research Report No. 19. Uxbridge, UK: HERG, Brunel University.

CANADA REVENUE AGENCY, July 1st, 2008-last update, charities listing [Homepage of Canada Revenue Agency], [Online]. Available: http://www.cra-arc.gc.ca/ebci/haip/srch/sec/SrchLogin-e;jsessionid=LH96L50ByvXMC9bWrns33tgyCPC2XwjsR6hTVqLpq8QnKs1YF3vh!39097939?searchType=Registered [July 23rd, 2008].

CANADIAN AGENCY FOR DRUGS AND TECHNOLOGIES IN HEALTH, June 28th 2008, 2008-last update, common drug review [Homepage of Canadian Agency for Drugs and Technologies in Health], [Online]. Available: http://cadth.ca/index.php/en/cdr [June 28th, 2008].

CANADIAN HEALTH SERVICES RESEARCH FOUNDATION, n.d.a. *Is research working for you?* Ottawa, Ontario: CHSRF.

CANADIAN HEALTH SERVICES RESEARCH FOUNDATION, n.d.b-last update, statement of institutional purpose [Homepage of CHSRF], [Online]. Available: http://www.chsrf.ca/about/do_statement_purpose_e.php [July 24th, 2008].

CANADIAN HEALTH SERVICES RESEARCH FOUNDATION, 2005. *Annual Report.* Ottawa, Ontario: CHSRF.

CANADIAN INSTITUTES FOR HEALTH RESEARCH, November 2007, 2007-last update, CIHR framework to measure the impact of health research: "the socio-economic returns of medical research workshop" [Homepage of Swedish Research Council], [Online]. Available: http://www.vr.se/download/18.5ec4ab911161ebf14dc80005426/CIHR+Borbey.pdf [July 22nd, 2008].

CANADIAN INSTITUTES FOR HEALTH RESEARCH, 2005. *Developing a CIHR Framework to Measure the Impact of Health Research.* Ottawa, Ontario: CIHR.

CANADIAN INSTITUTES FOR HEALTH RESEARCH, 2004. *Investing in Canada's Future: CIHR's Blueprint For Health Research and Innovation.* Ottawa; Ontario: CIHR.

CANADIAN INSTITUTES FOR HEALTH RESEARCH, 2001. *Canadian Institutes of Health Research: Part III – Report on Plans and Priorities; 2001-2002 Estimates.* Ottawa, Canada: Canadian Government Publishing (PWGSC).

CIHR, 2008. Payback categories. Ottawa, Canada: CIHR.

COUNCIL FOR MEDICAL SCIENCES, 2002. *The societal impact of applied health research: Towards a quality assessment system.* Amsterdam, the Netherlands: Royal Netherlands Academy of Arts and Sciences.

COUNCIL ON HEALTH RESEARCH FOR DEVELOPMENT (COHRED), 2006. *Supporting health research systems development in Latin America: Results of Latin America Regional Think Tank.* Record Paper 6. Geneva, Switzerland: COHRED.

DUSSAULT, G., DAVIS, J., GRUMAN, J. and THORNTON, S., 2007. *CHSRF at the crossroads - Capitalizing on success: Report of the International Review Panel To the Board of Trustees of the Canadian Health Services Research Foundation.* Ottawa, Ontario: CHSRF.

ELIAT, H., GOLANY, B. and SHTUB, A., 2006. Constructing and evaluating balanced portfolios of R&D projects with interactions: a DEA based methodology. *European Journal of Operational Research,* **172**(3), pp. 1018-1039.

ELLENBROEK, S., 2007. The experiences of the Leiden University Medical Center with its Societal Impact project and its potential for national implementation. Presentation to the Swedish Research Council Workshop on Economic Returns of Medical Research edn. http://www.vr.se/download/18.34261071168fe6a62080001006/Stefan+Ellenbroek+2.pdf: .

FUNDING FIRST, 2000. *Exceptional Returns: The economic value of America's investment in medical research.* New York, New York: The Lasker Foundation.

GARRETT-JONES, S., TURPIN, T. and WIXTED, B., 2004. *Science for Life - An Evaluation of New Zealand's Health Research Investment System based on International Benchmarks, Executive Summary.* New Zealand Ministry of Research, Science and Technology.

GERHARDUS, A. and DINTSIOS, C., 2005. The impact of HTA reports on health policy: a systematic review. *GMS Health Technology Assessment,* **1**, pp. Doc02.

GRANT, J., COTTRELL, R., CLUZEAU, F. and FAWCETT, G., 2000. Evaluating "payback" on biomedical research from papers cited in clinical guidelines: applied bibliometric study. *British Medical Journal,* **320**, pp. 1107-1111.

HANNEY, S., MUGFORD, M., GRANT, J. and BUXTON, M., 2005. Assessing the benefits of health research: Lessons fromresearch into the use of antenatal corticosteroids for the prevention of neonatal respiratory distress syndrome. *Social Science & Medicine,* **60**, pp. 937-947.

HANNEY, S.R., BUXTON, M., GREEN, C., COULSON, D. and RAFFERTY, J., 2007. *An assessment of the impact of the NHS Health Technology Assessment Programme.* 11 (53). Norwich, UK: HMSO.

HANNEY, S.R., FRAME, I., GRANT, J., GREEN, P. and BUXTON, M.J., 2003. *From Bench to Bedside: Tracing the Payback Forwards from Basic or Early Clinical Research - A Preliminary Exercise and Proposals for a Future Study.* HERG Research Report No 31. Uxbridge, UK: HERG, Brunel University.

HEALTH CHARITIES COALITION OF CANADA, 2007-last update, members of the health charities coalition of canada [Homepage of HCCC], [Online]. Available: http://www.healthcharities.ca/members.html [July 23rd, 2008].

HEART AND STROKE FOUNDATION OF CANADA, 2008-last update, its payback time: new international study to assess impact of heart and stroke research [Homepage of HSFC], [Online]. Available: http://www.heartandstroke.com/site/apps/nlnet/content2.aspx?c=ikIQLcMWJtE&b=3485819&ct=4512407 [July 24th, 2008].

HEART AND STROKE FOUNDATION OF ONTARIO, 5th December, 2007-last update, request for proposal: evaluation of heart and stroke foundation of Ontario's healthy weights area for investment in mission (HW AIM) initiative [Homepage of HSFC], [Online]. Available: http://www.hsf.ca/research/en/provincial-programs/hsfo-rfp-evaluation-of-hw-aim.html [July 24t, 2008].

INDUSTRY CANADA, 2008. *Science and Technology data - 2006.* Ottawa, Ontario: Government of Canada.

JORDAN, G.B., MOTE, J. and HAGE, J.T., 2006. The Balanced Scorecard for R&D: Incorporating Measures of Technical Progress, *Atlanta Conference on Science and Technology Policy, Atlanta, GA,* May 18-19 2006, .

KAMPER-JORGENSON, F., ARBER, S., BERKMAN, L., MACKENBACH, J., ROSENSTOCK, L. and TEPERI, J., 2004. *International Panel Evaluation of Swedish Public Health Research.* Stockholm, Sweden: Swedish Council for Working Life and Social Research; Swedish National Institute of Public Health.

KAPLAN, R.S. and NORTON, D.P., 1996. Using the balanced scorecard as a strategic management system. *Harvard Business Review,* **Jan – Feb**, pp. 75-85.

KAPLAN, R.S. and NORTON, D.P., 1992. The Balanced Scorecard - Measures that Drive Performance. *Harvard Business Review,* (January-February), pp. 71-79.

KARLSSON, S. and WADSKOG, D., 2007. *A bibliometric survey of Swedish scientific publications between 1982 and 2004.* Department for Research Policy Analysis, Swedish Research Council.

KAROLINSKA INSTITUTET, January 9th, 2008-last update, bibliometrics at karolinska institutet [Homepage of Karolinska Institutet], [Online]. Available: http://ki.se/ki/jsp/polopoly.jsp?l=en&d=1610&a=17742 [July 31st, 2008].

KING, D.A., 2004. The scientific impact of nations. *Nature,* **430**(6997), pp. 311-316.

KURUVILLA, S., MAYS, N., PLEASANT, A. and WALT, G., 2006. Describing the impact of health research: a Research Impact Framework. *BMC Health Services Research,* **6**(134),.

KWAN, P., JOHNSTON, J., FUNG, A.Y., CHONG, D.S., COLLINS, R.A. and LO, S.V., 2007. A systematic evaluation of payback of publicly funded health and health services research in Hong Kong. *BMC health services research,* **7**, pp. 121.

LAFOND, A., KLEINAU, E., SHAFRITZ, L., PRYSOR-JONES, S., MBODJ, F., TRAORE, B., DEMBELE, E., GUEYE, M., BOUARÉ, M. and SOW, C., 2001. *Utilisation des données sanitaires pour améliorer les prestations de services: Une approche d'auto-évaluation.* Washington, D.C.: Support for Analysis and Research in Africa (SARA).

LAFORTUNE, L., FARAND, L., MONDOU, I., SICOTTE, C. and BATTISTA, R., 2008. Assessing the performance of health technology assessment organizations: a framework. *International Journal of Technology Assessment in Health Care,* **24**(1), pp. 76-86.

LANDRY, R., AMARA, N. and LAMARI, M., 2001. Climbing the Ladder of Research Utilization: Evidence from Social Science Research. *Science Communication,* (396), pp. 422.

LAVIS, J., ROSS, S., MCLEOD, C. and GILDINER, A., 2003. Measuring the impact of health research. *Journal of health services research & policy,* **8**(3), pp. 165-170.

LIPPI, G., FRANCHINI, M. and GUIDI, G.C., 2008. Doping in competition or doping in sport? *British Medical Bulletin,* **86**(1), pp. 95-107.

MANNING, F.J., MCGEARY, M. and ESTABROOK, R., 2004. NIH Extramural Center Programs: Criteria for Initiation and Evaluation. Bethesda, Maryland: National Institutes of Health.

MCGUIRE, A. and RAIKOU, M., 2007. *Inferring the Value of Medical Research to the UK.* Working Paper No: 5/2007. London, UK: London School of Economics Health.

MEDICAL RESEARCH COUNCIL, 2008a-last update, about us - evaluation reports [Homepage of Medical Research Council], [Online]. Available: http://www.mrc.ac.uk/AboutUs/CorporatePublications/EvaluationReports/index.htm [July 30th, 2008].

MEDICAL RESEARCH COUNCIL, 2008b-last update, our research: recent achievements [Homepage of Medical Research Council], [Online]. Available: http://www.mrc.ac.uk/OurResearch/Impact/ReecentAdvances/index.htm [30th July 2008, 2008].

MODELL, S., 2004. Performance measurement myths in the public sector: A research note. *Financial Accountability & Management,* **20**(1), pp. 39-55.

MOED, H.F., 2005. Citation Analysis in Research Evaluation. Dordrecht, Netherlands: Springer.

MURPHY, K.M. and TOPEL, R., 1999. The Economic Value of Medical Research.

NASON, E., JANTA, B., HASTINGS, G., HANNEY, S., O'DRISCOLL, M. and WOODING, S., 2008. *Health Research – Making an Impact: The Economic and Social Benefits of HRB Funded Research.* Dublin: Ireland: HRB Ireland.

NATIONAL ACADEMY OF SCIENCES COMMITTEE ON SCIENCE, ENGINEERING AND PUBLIC POLICY, 1999. *Evaluating Federal Research Programmes: Research and the Government and Results Act.* Washington, D.C.: National Academy Press.

NATIONAL HEALTH AND MEDICAL RESEARCH COUNCIL, 2007. *NHMRC FINAL REPORT SUMMARY: Preliminary results from final reports received for NHMRC funded awards that ended between 2003 – 2005.* http://www.nhmrc.gov.au/funding/funded/results/_files/2003_05_final_pre_results.pdf: Government of Australia.

NATIONAL HEALTH AND MEDICAL RESEARCH COUNCIL, 2004. *NHMRC's Performance Measurement Framework 2003-2006.* http://www.nhmrc.gov.au/publications/synopses/pmf2006.htm: Australian Government.

NATIONAL INSTITUTE OF PUBLIC HEALTH - SWEDEN, 2003. *Research Department Activity plan 2003.* 2003:21. Stockholm, Sweden: National Institute of Public Health.

NELSEN, G., 2008. *Implementing Metrics Management for Improving Clinical Trials Performance.* http://www.b-eye-network.com/view/7981: Business Intelligence Network.

NEVILLE, D., GATES, K., MACDONALD, D., BARRON, M., TUCKER, S., COTTON, S., FARRELL, G., HOEKMAN, T., BORNSTEIN, S. and O'REILLY, S., 2003. *Toward an Evaluation Framework for Electronic Health Records Initiatives: A Proposal For an Evaluation Framework.* Ottawa, Canada: Health Canada.

OFFICE OF THE ASSISTANT SECRETARY FOR PLANNING AND EVALUATION, 2007. *Program Evaluation Activities of the U.S. Department of Health and Human Services: Performance Improvement 2007.* Washington, D.C.: Department of Health and Human Services.

OORTWIJN, W.J., HANNEY, S.R., LIGTVOET, A., HOORENS, S., WOODING, S., GRANT, J., BUXTON, M.J. and BOUTER, L.M., 2008. Assessing the impact of health technology assessment in the Netherlands. *International Journal of Technology Assessment in Health Care,* **24**(3), pp. 259-269.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, 2002. *Frascati Manual: Proposed Standard Practice for Surveys on Research and Experimental Development.* Paris, France: OECD Publications Service.

OSAMA, A., 2006. *Multi-Attribute Strategy and Performance Architectures in R&D: The Case of The Balanced Scorecard*, RAND Corp. PRGS.

PEACH, I. and MARSHALL, J., 2008. *The Economic Benefits of Health Research in Saskatchewan.* Regina, Saskatchewan: Saskatchewan Institute of Public Policy.

PROFETTO-MCGRATH, J., HESKETH, K.L., LANG, S. and ESTABROOKS, C.A., 2003. A Study of Critical Thinking and Research Utilization Among Nurses. *Western Journal of Nursing Research,* **25**(3), pp. 322-337.

REED, J.H. and JORDAN, G., 2007. Using systems theory and logic models to define integrated outcomes and performance measures in multi-program settings. *Research Evaluation,* **16**(3), pp. 169-181.

RESEARCH EVALUATION AND POLICY PROJECT, August 2nd, 2008-last update, about REPP [Homepage of Australian National University], [Online]. Available: http://repp.anu.edu.au/ [July 31st, 2008].

RESEARCH EVALUATION AND POLICY PROJECT, 2005. *Quantitative Indicators for Research Assessment - A Literature Review.* Review No. 05/1. Canberra, ACT: Research School of Social Sciences, Australian National University.

ROGERS, C.R. and FREIBERG, H.J., 1994. Freedom to learn. 3rd edn. New York, USA: Macmillan College Publishing Company.

ROSENBERG, L.E., 2002. Exceptional economic returns on investments in medical research. *The Medical journal of Australia,* **177**(7), pp. 368-371.

RUEGG, R. and FELLER, I., 2003. *A Toolkit for Evaluating Public R&D Investment: Models, Methods and Findings from ATP's First Decade.* NIST GCR 03-857.

RX&D, 2007. *The Pharmaceutical Industry in Canada.* September. Ottawa, Ontario: Rx&D.

RX&D, 2006a. *Canada's research based pharmaceutical companies: Improving and saving lives.* Ottawa, Ontario: Rx&D.

RX&D, January 23rd, 2006b-last update, did you know? our community [Homepage of Rx&D], [Online]. Available: http://www.canadapharma.org/DidYouKnow/index_e.html [July 27th, 2008].

SANDSTRÖM, U., 2008. *Combining Curriculum Vitae and Bibliometric Analysis: Mobility, Gender and Research Performance.* Linkoping, Sweden: Linkoping University.

SASKATCHEWAN HEALTH, 2004. *Health Research Strategy: Saskatchewan. Healthy People. A Healthy Province.* Regina, Saskatchewan: Saskatchewan Health.

SASKATCHEWAN HEALTH RESEARCH FOUNDATION, January 2007. *Evaluating Saskatchewan's Health Research Strategy: An Evaluation Framework.*

SCARPA, T., 2007. *Funding Biomedical Research In the U.S.A.: The Strategic National Importance of NIH Peer Review.* Sigtuna, Sweden: Economic returns of medical research meeting.

SCIQUEST, n.d.-last update, SciQuest: scientific consultancy agency for research and higher education policy [Homepage of SciQuest], [Online]. Available: http://www.xs4all.nl/~jbspaa/ [July 30th, 2008].

STATISTICS CANADA, 2008. *Science Statistics May 2008 edition.* Catalogue no. 88-001-X. Ottawa, Ontario: Statistics Canada.

THE KIDNEY FOUNDATION, n.d.-last update, 2007 KFOC medal for research excellence [Homepage of The Kidney Foundation], [Online]. Available: http://www.kidney.ca/page.asp?intNodeID=35458 [24th July, 2008].

TIESSEN, J., 2007. *Health and Medical Research in Sweden: Observatory on Health Research Systems.* DB-533-DH. Cambridge, UK: RAND Europe.

TROSTLE, J., BRONFMANN, M. and LANGER, A., 1999. How do researchers influence decision makers? Case studies of Mexican Policies. *Health Policy and Planning,* **14**, pp. 103-114.

UK EVALUATION FORUM, 2006. *Medical research: assessing the benefits to society.* London, UK: UK Evaluation Forum.

UNESCO INSTITUTE FOR STATISTICS, 2005. *What do bibliometric indicators tell us about world scientific output?* Issue No. 2. Montreal, Quebec: UNESCO Institute for Statistics.

UNIVERSITY HEALTH NETWORK, n.d. 2008a-last update, accountability to our patients: how we measure up - the balanced scorecard [Homepage of University Health Network], [Online]. Available: http://www.uhn.ca/Patients_&_Visitors/wait_times_for_care/index.asp [September 4th, 2008].

UNIVERSITY HEALTH NETWORK, n.d. 2008b-last update, performance measurement [Homepage of University Health Network], [Online]. Available: http://www.uhn.ca/About_UHN/corporate_info/oppm/site/corporate/corporate.asp [September 4th, 2008].

UNIVERSITY HEALTH NETWORK, n.d. , understanding UHN strategy and the balanced scorecard [Homepage of University Health Network], [Online]. Available: http://www.uhn.ca/Patients_&_Visitors/wait_times_for_care/docs/IntroductiontotheBalancedScorecard%20%2008_09.pdf [September 4th, 2008].

VAN ARK, G., 2007. Societal impact evaluation of research groups: The communication metaphor. Presentation to the Swedish Research Council Workshop on Economic Returns of Medical Research edn. http://www.vr.se/download/18.34261071168fe6a62080001004/ZonMw+van+Ark+2.pdf: .

VAN LEEUWEN, T.N. and GRANT, J., 2007. *Bibliometric analysis of highly cited publications of health research in England, 1995-2004.* WR-368-DH. Cambridge, UK: RAND Europe.

VETENSKAPSRÅDET, 2004. *The Medical Research Situation in Sweden.* Stockholm, Sweden: Vetenskapsrådet.

WALT, G. and GILSON, L., 1994. Reforming the health sector in developing countries: the central role of policy analysis. *Health Policy and Planning,* **9**(4), pp. 353-370.

WEBSTER, B.M., LEWISON, G. and ROWLANDS, I., 2003. *Mapping the Landscape II: Biomedical Research in the UK, 1989-2002.* London, UK: School of Informatics, City University.

WEISS, A.P., 2007. Measuring the Impact of Medical Research: Moving From Outputs to Outcomes. *American Journal of Psychiatry,* **164**(2), pp. 206-214.

WELLCOME TRUST, 2005. *Strategic Plan 2005-2010: Making a Difference.* London, UK: Wellcome Trust.

WELLCOME TRUST, a-last update, how we make a difference [Homepage of Wellcome Trust], [Online]. Available: http://www.wellcome.ac.uk/Achievements-and-Impact/How-we-make-a-difference/index.htm [July 30th, 2008].

WELLCOME TRUST, b. Public Engagement: Evaluation guidelines for applicants. Guidelines edn. http://www.wellcome.ac.uk/stellent/groups/corporatesite/@msh_grants/documents/guidance/wtx035088.doc.: Wellcome Trust.

WOODING, S., 2008a. MRC evaluation work - ongoing. Cambridge, UK: RAND Europe.

WOODING, S., 2008b. Utilising a payback survey to assess Arthritis Research Council (arc) funding - ongoing research. Cambridge, UK: RAND Europe.

WOODING, S., 2006. *Reviewing the returns of research: capturing payback from funding by the Arthritis Research Campaign.* RB-9199-ARC. RAND Europe.

WOODING, S., ANTON, S., GRANT, J., HANNEY, S., HOORENS, S., LIERENS, A., SHERGOLD, M. and VENEMA, M., 2004. *The returns from arthritis research - Volume 2: Case Studies.* TR-176-ARC. Cambridge, UK: RAND Europe.

WOODING, S., HANNEY, S., BUXTON, M. and GRANT, J., 2005. Payback arising from research funding: evaluation of the Arthritis Research Campaign. *Rheumatology (Oxford),* **44**, pp. 1145-56.

WOODING, S., HANNEY, S., BUXTON, M. and GRANT, J., 2004. *The Returns from Arthritis Research Volume 1: Approach, Analysis and Recommendations.* MG-251-ARC. Cambridge, UK: RAND Europe.

WOODING, S., NASON, E., KLAUTZER, L., RUBIN, J., HANNEY, S. and GRANT, J., 2007. *Policy and practice impacts of research funded by the Economic and Social Research Council: A case study of the Future of Work programme, approach and analysis.* Cambridge, UK: RAND Europe.

Appendix D: Issues for research evaluation

Particular problems always arise when attempting to evaluate research impacts: attribution, the counterfactual; threats to evaluation validity; and time lags to research impacts. Since these issues consistently arise, it can be assumed that none has been fully solved. In this appendix we highlight these four issues and the ways that they can be addressed in order to improve the quality, reliability and robust nature of the evaluation in hand.

Attribution

Devising and developing an evaluation framework that can accurately represent the full range of impacts arising from research (in health or otherwise) can be a difficult task. However, attempting to catalogue impacts is relatively straightforward when compared with deciding what the contributing factors to those impacts were. The problem of attribution of impacts is not one particularly to research, but it is one that must be addressed if one truly wishes to understand the impacts of research funding. In this chapter we will provide some of the background to the attribution problem, how different aspects of attribution pose different problems, how other people evaluating the impacts of health research have addressed the attribution issue, and finally, how we propose to address attribution within our own framework and metrics.

Clearly it is not possible to begin any discussion of attribution without really understanding what we mean by the word itself:

> *'Attribution: The ascription of a causal link between observed (or expected to be observed) changes and a specific intervention.'*

Source: (Organisation for Economic Cooperation and Development (OECD) 2002)

Although the above is a good working definition, it is noticeable that the OECD also provides an additional note that attribution more specifically relates to the extent that any outcome can be ascribed to the intervention[113] whilst taking into account all the other factors that may have been involved in the outcome.

It is in these 'other factors' that the great difficulty of attribution occurs. The prevalence of factors influencing an outcome is often associated with the distance from the intervention itself (Figure 45)(Smutylo 2001). Therefore, in the case of health research, final outcomes of health change and economic impacts are often those most difficult to link to the research itself; whereas additions to the knowledge pool such as publications and books can be relatively easily tied to specific funding (although this is not always the case, for example if a book is the result of a body of work funded by a variety of different health R&D funders).

---

[113] For our purposes, an intervention here represents a piece of health R&D or funding for health R&D.
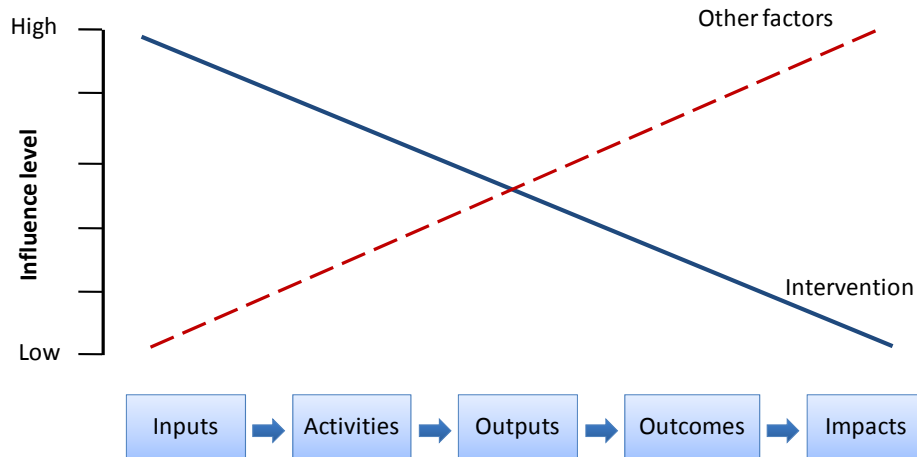
Figure 45. Relative influence of an intervention and other factors in the logic chain (adapted from Smutylo 2001)

Attribution can be an issue at a number of different levels. These can usefully be split into four main levels.

At the highest level, impacts observed need to be attributed to research in general (i.e. they must be attributable to research at least partially).

At a level below research, is attributing specifically to health research. As an example of this is the often cited case of cardiovascular disease, where several research projects (including the 'Exceptional Returns' studies) have attributed one third of the health impacts and cost savings to the healthcare system to research (Cutler, Kadiyala 1999, Wisconsin Association for Biomedical Research & Education 1995). This is based on the contribution of new techniques and drugs, so is likely to be something of an underestimate – acknowledged in the Access Economics report, where they bump up the percentage attributable to research to 50% (Access Economics Australia Economic Consulting 2003). This percentage links the impact on health to any research at all – not even specific to that done within a country.

This links nicely to the third level of attribution, that of research from a specific country. For our purposes this is relevant to health research specifically from Canada, but one could also consider research from a particular country as a second level of attribution (before specifically looking at health research). A good example of this issue in Canada has been the introduction of smoking bans in public places across Provinces. These policy decisions used extensive public health research to support them, but a large proportion of that research came from international sources and evaluations of smoking bans put in place around the world (Schmidt 2007, Glantz, Parmley 2001). It is very difficult to identify a percentage attribution to Canadian health research that informed this decision since there are so many different health research inputs.

Finally, it would be very useful for evaluations of health research funding to be able to link specific research programs, projects, or funding streams to particular impacts. This would allow identification of the attribution of specific research on impacts. However, this is also the most difficult level at which to attribute since, as shown above, there are many factors involved in most impacts associated with health research. There are examples of impacts that can be specifically linked to funding or programs, such as the changes in survivorship for heart attacks in Ireland based on health services research that suggested changes to the speed of treatment in hospital emergency rooms (Nason, Janta et al. 2008); but these examples are few and far between.

The problem of attributing outcomes is not new to research, and is not solely an issue for R&D outcomes. In any evaluation, being able to clearly understand the way the inputs have affected the

outcomes is key to learning anything from the evaluation itself. It is interesting to note that international development and agriculture seem to be the fields most synonymous with theoretical research into attribution to research and/or interventions. This is likely because, particularly in the case of international development, the aim of an intervention is to remove any need for the intervention, thus understanding how much of the change in impacts is due to the intervention itself is crucial.

By having a logic model of the way research progresses to outcomes, it is easier to begin to attribute impacts to research findings (Program Evaluation Branch, Office of the Comptroller General 1991). Logic models allow an identification of the narrative of any impact identified. By using that narrative and tracing back through the model it is possible to identify the likely contributing factors to any impact.

In a 1999 paper by the Office of the Auditor General in Canada (Mayne 1999), John Mayne suggests using a technique called 'Contribution Analysis'. In its most basic terms contribution analysis is a set of steps that any evaluation should go through in order to try to identify the contribution to impacts made by an intervention (Figure 46), a process similar to tracing within a logic model of research impacts.
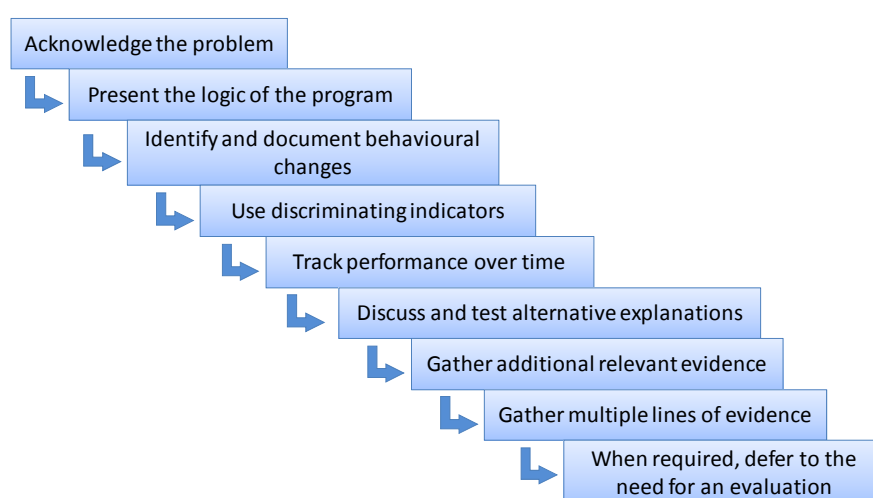
Figure 46. The sectors of a consultation analysis (Mayne 1999)

Perhaps the key aspect within the contribution analysis technique though is the idea of tracking performance over time. By effectively following the performance of the intervention and any ongoing externalities, it is possible to arrive at a better understanding of how impact changes over time are related to the intervention itself (be that directly, indirectly or unexpectedly).

To make a judgement on the level of attribution that any contributing factor has had it is necessary to collect as much qualitative and quantitative information as possible, usually through in depth case studies of specific research funding programs or projects. This can help to identify how research has progressed through various actors towards a final outcome (be that a health improvement or a change in policy), but it relies upon a single key assumption that actors in the change will be able to say how much a particular piece of research has affected their thinking. For example, in work for the Economic and Social Research Council in the UK, it was discovered that policy makers tended to know particular researchers rather than particular research outputs, and therefore the policy makers' thinking tended to be informed by a body of research rather than specific findings (Wooding, Nason et al. 2007). In the recent work for the UK evaluation forum (Buxton, Hanney et al. 2008), the researchers used a large volume of quantitative data on inputs, processes and pathways to impacts to try and attribute economic impacts to health research in cardiovascular disease and mental health (their two case studies of the economic return on medical research funding). Using this large amount of quantitative data required the

research team to make a number of assumptions, and this is a common issue for attribution (the concept of 'heroic assumptions'). It has been the main criticism of econometric studies of the return on health research investment (see Shiel and Di Ruggiero's paper in Appendix A for further details of these criticisms).

Case studies come with the potential problem of 'representativeness', since limited evaluation resources often means only a small number of case studies can be performed. This can be resolved by creating rigorous sampling techniques that allow the case studies selected to form a representative sample of the funding under evaluation. It should be noted, however, that in evaluations of research impact, identifying studies that have been high or low impact (in order to represent possible outcomes of research) is difficult prior to the evaluation.[114] Work in the UK by Brunel University on the Payback Model has been using case study analysis for some time, addressing different research funders and funding systems in a highly qualitative manner. As such, they have been building up a significant library of case study data which it would be very interesting to perform a meta-analysis on (Hanney, pers. com.).

Regression analyses (also known as causal models: (Program Evaluation Branch, Office of the Comptroller General 1991)), a statistical technique for modeling and analysis of numerical data consisting of values of a dependent variable (response variable) and of one or more independent variables (explanatory variables), can also be used to understand the relationships between research funding and impacts. It is hard within the data base of a single nation to prove a relationship between a putative input ('government funding on health research', say, or 'private funding on health research') and the desired outputs (improvements in the mortality and morbidity rates). But multivariate regressions across different but comparable nations operating different policies may allow greater signal-to-noise ratios to emerge, thus allowing us to establish empirically that other inputs such as the government or private funding of health research will impact on the health of the nation.

Whilst regression studies of the impact of health research on changes to mortality and morbidity have not been performed, other complex regression analyses have been pursued. These include a multivariate growth regression for 21 OECD countries using 27 years worth of data that investigated the explanatory variables for changes in GDP (Organisation for Economic Co-Operation and Development 2003). Interestingly, this study showed that there were differences between the effects of publicly and privately funded R&D on GDP and that the two funding sources did not act complimentarily (although this was all R&D, not just health related, where other evidence suggests that public and private research interact positively (Cumbers, Birch 2006); (Congressional Budget Office 2006); (Joint Economic Committee 2000)). The authors of the OECD report emphasised a key issue with using multivariate analyses; that they depend on good data and a sound understanding of the factors involved in the response variable. With better data and understanding of the causes of health impacts, it should be an ambition to extend the sort of multivariate analysis the OECD used for GDP per capita into mortality and morbidity studies, to try to establish the role of health R&D funding.

Most of the solutions to the attribution issue have concerned the over-attribution of impacts to research results (since most evaluations come from those wishing to show the impacts of their research). Over-attributing can lead to double-counting of impacts since multiple research factors will all be considered to have been responsible for the impact. Whilst this is the most common risk with attribution in research evaluation, there is also a real risk of under-attribution in some situations. For example, whilst

---

[114] In selecting case studies for the ongoing 'Project Retrosight' (Heart and Stroke Foundation of Canada 2008), the researchers used a short survey to understand whether potential case studies would be high or low impact. This survey was based on outcomes from research such as publications, training of researchers, changes to policy and changes to health care.

evaluations commonly identify the explicit outputs and outcomes from research such as product development, tacit research outputs such as interactions with decision makers are often overlooked, implying that the research impact on that decision maker is not identified. Being able to address under-attribution also uses the same methods: increasing volumes of data, better narratives of how research has impacts and a more nuanced understanding of the contributing factors in impacts. By improving the methods for understanding attribution in general we should be able to progressively home in on the 'real' value of research's role in impacts and deal with both over-attribution and under-attribution.

Due to the difficulties of solid attribution, some evaluations have accepted that what should be addressed is contribution. For example the use of outcome mapping in international development has decided to utilise contribution rather than try to understand full attribution since being involved in the outcome is a 'good enough' measure in some situations (especially considering most of the methods mentioned above are either resource intensive or require external expertise to perform) (Rosenberg 2002).

The counterfactual

Attributing the actual impacts of research can be tricky, but at least represents a known impact. An added complication arises for evaluations in understanding what would have happened without the intervention (research) under evaluation – the counterfactual. This counterfactual situation is most commonly identified by using a comparison (or control) group for the evaluation (Earl, Carden 2002). The gold standard for evaluation with a comparison group is the randomised control trial (Organisation for Economic Co-Operation and Development n.d.), but this is often not possible with research funding. In these cases, quasi-experimental designs allow comparisons across different groups involved in the evaluation, but do not involve random assignment to groups. Quasi-experimental designs for evaluation still provide an idea of a counterfactual, but do not provide the exact situation that might have occurred without funding (or with funding provided elsewhere). For any evaluation design, identifying baseline measures and context is important in understanding what any counterfactual might have looked like. Having a framework that can understand the different external contextual factors that may have been involved in impacts makes understanding the counterfactual easier.

Internal and external threats to evaluation validity

An understanding of the likely threats to validity for any evaluation (Table 3) is important to avoid falling into common problems. These threats can be internal to the evaluation undertaken, i.e. that the problems are ones that could undermine the findings of the evaluation itself; or external to the evaluation, i.e. that the evaluation itself may not be generalizable to other situations.

Internal threats to the validity of the evaluation relate strongly to the counterfactual situation and contextual factors surrounding the impacts identified. They include not taking into account the history prior to the intervention (in our case health research funding); changes that occur during the research process itself (such as research performed elsewhere feeding into the project under evaluation); selection bias in the evaluation (for example around choosing case studies); and measurement issues (such as different interpretations of interview protocols).

External threats are more closely linked to the context of the evaluated program or funding. They include whether the evaluated program is representative of the funding it is being generalised to (something that careful case study selection protocols can help to solve); that the evaluated program or funding does not take place in a context that is representative of other funding (e.g. the evaluated research may have taken place in a multi-disciplinary unit, whereas other funding has gone to uni-

disciplinary research departments); and whether the history and time of the program or research is representative of the time in which impacts are occurring (Duflo, Kremer 2003).

Table 3. Common internal and external threats to validity of an evaluation (Program Evaluation Branch, Office of the Comptroller General 1991)

Internal threats to validity	External threats to validity
• History of participants prior to the intervention. • Maturation of participants during the program. • Mortality of participants in the intervention (mortality hear meaning drop-out rate). • Selection bias in those within the intervention. • Regression artefacts – all groups eventually progress towards the mean regardless of whether the intervention has benefited them or not. • Measurement issues – different groups may be measured differently (e.g. using the same interview protocol in different ways). • Diffusion of the intervention to those not expected to be affected by it.	• Selection of participants for any intervention is not representative of the population. • Setting for the intervention may not be representative • History prior to the intervention may not be representative of future conditions

Time lags to research impacts

Time-lags exist between research funding and impacts and trying to identify what the time lags are for different types of research can be very difficult. For example we know that health services research is more likely to have a fast effect on cost effectiveness of healthcare, whilst basic science may take a very long time to truly impact on health; but we do not know the timescales for any piece of research in these disciplines to have impacts. Impacts can also persist for different lengths of time or indeed change from being positive impacts to negative ones depending upon the findings from other research projects. As an example, recent studies into Hormone Replacement Therapy (HRT) have shown new dangers associated with long-term treatment that were not known previously, a situation that shows the importance of understanding time lags in addressing impacts.

The most obvious way to understand time lags for different types of research is to collect data on the lengths of time it takes for research to inform specific impacts. This is the approach taken to identifying the time lag between research performed in cardiovascular disease and the changes to health care practice in the UK. In the recent work for the UK Evaluation Forum, researchers identified the time from research being published to its inclusion in clinical guidelines for specific conditions as a way to elaborate the time lag to a clinical (and therefore assumed health) impact (Program Evaluation Branch, Office of the Comptroller General 1991). Other studies have identified the time lags for particular aspects of impact: for example the time from publishing to having a measurable citation impact is considered to be around 3-5 years (Buxton, Hanney et al. 2008); while the time lag estimated for the findings of high-quality randomized controlled trials to become standard clinical practice is up to 17 years (see Lewis et al – Appendix A).

In estimating lags, it is very difficult to match up the time to effect and the persistence of the effect (something that will also affect the size of impact). If we take a hypothetical example, we can identify a piece of population health research that has 5 impacts, each of which occurs at a different time. We can

create a distribution of impacts by plotting the size of the impact using a standardized measure, their time of occurrence and their duration. For impacts that have a finite lifespan we can relatively simply model this distribution, for those impacts that persist, the distribution becomes more problematic.

Some evaluations, mainly of agricultural research, have attempted to use an econometric technique called 'infinite-lag modelling' to take these persistence effects into account (Van Leeuwen, Grant 2007). Infinite lag modelling uses weighted impacts (parameters in the model) based on the impact and its distance from the intervention. This is done to account for an infinite number impacts or duration of time lag to the impact (Alston, Pardey 2001). Infinite lag modelling requires large amounts of data, such as long term input funding data and observational data on outcomes and their durations and intensities, to model the potential different time lags that exist for different impacts. Considering that evaluating the impacts of health research is a relatively new approach, this level of data is not available for health; making infinite lag models currently unsuitable to address time lag issues.

References

ACCESS ECONOMICS AUSTRALIA ECONOMIC CONSULTING, 2003. Exceptional Returns: the value of investing in health R&D in Australia.

ALSTON, J.M. and PARDEY, P.G., 2001. Attribution and other problems in assessing the returns to agricultural R&D. *Agricultural Economics,* **25**, pp. 141-152.

BECK, N., 2004. Lag Structure. In: M.S. LEWIS-BECK, A. BRYMAN and T. FUTING LIAO, eds, *The Sage Encyclopedia of Social Science Research Methods: Volume two.* London, UK: Sage publications, pp. 546-547.

BUXTON, M., HANNEY, S., MORRIS, S., SUNDMACHER, L., MESTRE-FERRANDIZ, J., GARAU, M., SUSSEX, J., GRANT, J., WOODING, S., NASON, E., ISMAIL, S. and KAPUR, S., 2008. Medical Research – What's it worth? Estimating the economic benefits from medical research in the UK. Report to the UK Evaluation Forum edn. London, UK: .

CONGRESSIONAL BUDGET OFFICE, 2006. *Research and Development in the Pharmaceutical Industry.* Washington, D.C.: Congressional Budget Office.

CUMBERS, A. and BIRCH, K., 2006. *Public Sector Spending and Regional Economic Development: Crowding Out or Adding Value?* Glasgow, Scotland: Unison Scotland.

CUTLER, D.M. and KADIYALA, S., 1999. *The Economics of Better Health: The Case of Cardiovascular Disease.* New York, New York: The Lasker Foundation.

DUFLO, E. and KREMER, M., 2003. *Use of Randomization in the Evaluation of Development Effectiveness.* MIT, Cambridge, MA: World Bank Operations Evaluation Department (OED) Conference on Evaluation and Development Effectiveness.

EARL, S. and CARDEN, F., 2002. Learning from complexity: the International Development Research Centre's experience with Outcome Mapping. *Development in Practice,* **12**(3-4), pp. 518-524.

GLANTZ, S.A. and PARMLEY, W.W., 2001. Even a Little Secondhand Smoke Is Dangerous. *The Journal of the American Medical Association,* **286**(4), pp. 462-463.

HEART AND STROKE FOUNDATION OF CANADA, 2008-last update, its payback time: new international study to assess impact of heart and stroke research [Homepage of HSFC], [Online]. Available: http://www.heartandstroke.com/site/apps/nlnet/content2.aspx?c=ikIQLcMWJtE&b=3485819&ct=4512407 [July 24th, 2008].

JOINT ECONOMIC COMMITTEE, 2000. *The benefits of medical research and the role of the NIH.* Washington, D.C.: US Senate.

MAYNE, J., 1999. *Addressing Attribution Through Contribution Analysis: Using Performance Measures Sensibly.* Ottawa, Ontario: Officer of the Auditor General of Canada.

NASON, E., JANTA, B., HASTINGS, G., HANNEY, S., O'DRISCOLL, M. and WOODING, S., 2008. *Health Research – Making an Impact: The Economic and Social Benefits of HRB Funded Research.* Dublin: Ireland: HRB Ireland.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, n.d. *Outline of Principles of Impact Evaluation.* http://www.oecd.org/dataoecd/46/16/37671602.pdf: OECD.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, 2003. *The Sources of Economic Growth in OECD Countries.* Paris, France: OECD.

ORGANISATION FOR ECONOMIC COOPERATION AND DEVELOPMENT (OECD), 2002-last update, glossary of key terms in evaluation and results based management. Available: http://www.oecd.org/dataoecd/29/21/2754804.pdf.

PROGRAM EVALUATION BRANCH, OFFICE OF THE COMPTROLLER GENERAL, 1991. Program Evaluation Methods: Measurement and Attribution of Program Results. Third edn. Ottawa, Ontario: Public Affairs Branch, Treasury Board of Canada.

ROSENBERG, L.E., 2002. Exceptional economic returns on investments in medical research. *The Medical journal of Australia,* **177**(7), pp. 368-371.

SCHMIDT, C.W., 2007. A Change in the Air: Smoking Bans Gain Momentum Worldwide. *Environ Health Perspect.,* **115**(8), pp. A412-A415.

SMUTYLO, T., 2001. *Crouching Impact, Hidden Attribution: Overcoming Threats to Learning in Development Programs.* Ottawa, Ontario: Evaluation Unit, International Development Research Centre.

VAN LEEUWEN, T.N. and GRANT, J., 2007. *Bibliometric analysis of highly cited publications of health research in England, 1995-2004.* WR-368-DH. Cambridge, UK: RAND Europe.

WISCONSIN ASSOCIATION FOR BIOMEDICAL RESEARCH & EDUCATION, 1995. *Bioscience Research, Development & Industry: Impact On Health & Economic Growth in Wisconsin.* Available from http://www.wabre.org: WABRE.

WOODING, S., NASON, E., KLAUTZER, L., RUBIN, J., HANNEY, S. and GRANT, J., 2007. *Policy and practice impacts of research funded by the Economic and Social Research Council: A case study of the Future of Work programme, approach and analysis.* Cambridge, UK: RAND Europe.

Appendix E: Indicators

Frameworks can suggest what should be collected but data quality and availability are key to producing a robust and reliable evaluation. Many indicators and metrics are already in use around the world for health research, and here we present an overview of the most commonly used indicators in health evaluations and identify where most value can be added by identifying new indicators or improving the range or quality of existing indicators. For the sake of structure, we organise the indicators according to what aspect of impact they measure using the payback framework impact categories: Knowledge; Capacity Building; Informing Decision Making; Health; and Economic.[115] We do not present specific indicators that would only work for a single organization here (such as "Extent to which Institutes have appropriately influenced the research, policy and/or practice agendas in their communities" which is one of CIHR's specific indicators, but not one that can be translated to another context (Canadian Institutes for Health Research 2005)) since these are specialized and we wish to showcase generic indicators that can be picked up by any organization wishing to evaluate.

Advancing Knowledge Indicators

Knowledge gain is the most accessible category for indicators, albeit still problematic, because the outputs of research (usually publications) are closer to the research itself so can be counted and attributed relatively straightforwardly (Wells, Whitworth 2007). Australia has been particularly active in identifying knowledge metrics and indicators for research assessment, particularly in relation to the use of different bibliometric indicators. In their 2006 literature review of quantitative metrics for research assessment (Research Evaluation and Policy Project 2005), the Research Evaluation and Policy Project at the Australia's National University identified metrics that are useful and used to determine knowledge production. In Canada, CIHR use this category to include new discoveries and breakthroughs from health research, and contributions to the scientific literature (CIHR 2008). In our reworking of the category, there are four sub-categories: activity measures; quality measures; outreach measures; and structural measures.

Activity

Activity measures are those that identify the output from research funding. They collect any knowledge impact that occurs and make no differentiation between what quality levels they are, what stakeholder they were designated for or how they classify in terms of the greater output from all R&D in the country. It must be stated that counts of publications as an individual indicator are not useful, since they can drive behaviour away from quality research outputs, but with quality measures can be a useful indicator (Butler 2002). They are generally counts of outputs and in our classification fall into five further sub-groups:

a. Number of peer-reviewed publications

As the main indicator used to identify the activity output of health researchers, the number of peer-reviewed publications is a well established indicator in use around the world (Research Evaluation and Policy Project 2005, Research Evaluation and Policy Project 2005, Moed 2005a, Wooding 2008). The pure count of publications can also be normalized by a number of different factors to provide a more

---

nuanced view of the activity – for example per researcher; by amount of funding received; per annum; by discipline of research; by degree of specialization (Research Evaluation and Policy Project 2008).

b. Publication share

By identifying the output of publications by individuals, it is also possible to see how productive they are related to their peers in the world. This is basically identifying the share of the world's publications an individual (or collection of individuals such as a university or province) has (Research Evaluation and Policy Project 2005)(Research Evaluation and Policy Project 2008). Normalizing by research field provides a slightly more useful measure of output compared with the world since some disciplines produce a very large number of publications whilst others don't (molecular biology vs. health informatics for example).

c. Number of non-peer reviewed outputs

Not all areas of health research have peer-reviewed journal publications as their major output, instead publishing books, chapters of books, presenting at conferences, giving seminars or writing for trade or professional magazines (Wooding 2008). It is also true that as technology moves on, research outputs are being published on the internet rather than in traditional print formats. This means research findings can be published on personal or institution websites, through advocacy blogs, or via podcast.

d. Data indicators

The outputs of research can also be data as well as publications and this data can be submitted a variety of places. This includes international databases such as GenBank and the Entrez Protein (NCBI 2008b, NCBI 2008a), national databases such as the National Institute of Science and Technology (NIST) databases in the USA (National Institute of Standards and Technology 2008), or clinical trials databases such as the Clinical Trials Portal (Rx&D 2006a, Moher, Bernstein 2004). It is also possible to submit participation data on clinical trials, such as the number of recruits involved and the success rates for recruiting.

e. Technometric indicators

These represent the patents that are produced as an output of research, since they identify a different output to simple data or open publications data, rather representing the innovation output. Measurements of the number of patents produced are a common indicator in use in health research evaluation, including global comparisons through the OECD (Organisation for Economic Co-Operation and Development 2007).

<u>Quality</u>

Measuring the activity provides information on how productive research is, but what is required by funders is to produce the highest quality research outputs possible. For example, the CIHR mandate requires research excellence (Canadian Institutes of Health Research 2007). This can be identified in a number of ways.

a. 'Quality' normalized publication numbers

Journal Impact Factors (JIFs) represent a well known measure of publication 'quality'. They are calculated by dividing the number of current year citations to the source items published in that journal during the previous two years. This provides a ratio of citations to publications for the journal based on its recent articles (Institute for Scientific Information n.d.). However JIF has the drawback of relating only to the average citations for the journal, not identifying the quality of the individual publication. This problem also arises when normalizing using the mean number of journal citations or journal impact classes (Research Evaluation and Policy Project 2005). However, JIF is an understood concept and is easy to access (since it is publicly available), making it useful in starting to understand impact without having

to perform a full analysis. Publications can also be compared for their presence in 'high quality' journals or high quality outlets such as Cochrane Reviews, something that is being developed for books through identifying 'high quality publishers'. This concept can be stretched to include publications in outlets that target a specific stakeholder group, for example a journal set that is read by health practitioners, to help identify research that informs practice.

b. Citation measures

To really identify the impact of an individual publication, it is necessary to look at the number of citations it receives. Citation indicators have been well covered in the literature and their use is becoming more prevalent now in evaluation and in funding decision making (Vetenskapsrådet 2004, Van Leeuwen, Grant 2007, Nason, Grant et al. 2007, Tijssen, Van Leeuwen 2006) . Citations per publication are highly dependent upon the field of research the publication is in; therefore measures should always be normalized for field – becoming relative citation impact. Citations can be normalized by author, country or research funding received, can include or exclude self-citations, and can be weighted per citation (depending on where the citation occurs) (Research Evaluation and Policy Project 2005).

The two most common currently used quality indicators are the CWTS 'Crown Indicator', which uses the citations per publication average for an author or group and divides it by the average number of citations per paper for every publication in the world in the same research field, and Highly Cited Publication (HCP) counts, which identify how many publications from an author or group are amongst the most highly cited in the world in the same field (commonly used are top 10%, 5%, 2% and 1%). Both these measures normalize for field and compare to the rest of the world. The crown indicator identifies how good a group is on average compared to the world, and the HCP measure how often a group can produce very high quality publications.

Also worth mentioning in citation measures are the H-Index and its derivatives, and the ESI Hot Papers. The H-Index uses all the publications by an individual and finds the largest integer that represents the number of citations to papers and the papers that have that number or more citations (Van Raan 2006). For example, if a researcher has 100 publications, of which 25 have 25 or more citations, then their H-Index value would be 25. The H-index currently does not take into account research field or where the papers are published, but there is no reason why it shouldn't account for those and variations on the index are growing in number to make it more sophisticated (Bornmann, Mutz et al. 2008). The ESI Hot Papers measure identifies the top 0.01% worlds cited papers over a two month window, showing what the most cited work is currently (Essential Science Indicators 2008). This is interesting in identifying what research is in fashion, but does not work as a long-term evaluative tool.

c. Download numbers

For those publications that are not in journals in ISI but are published on the internet, for example publications from a university series on health research evaluation (Health Economics Research Group 2008), citations cannot be identified. Downloads of reports provide an alternative measure of impact, based on the assumption that downloads provide a similar measure to citations, an assumption that is currently being addressed in the bibliometric community (Moed 2005b). Ideally, a method for directly comparing citations and downloads will be identified, allowing direct comparisons of the two measures. Downloads of data can also be measured, but are more difficult to link to quality since they are not cited in research reports or publications.

d. Peer-review

In the USA, NIH rely exclusively on peer-review, although this is at the funding rather than evaluation stage (Scarpa 2007). This highlights the place peer-review has in identifying research quality, something

that citations can provide information on, but not replace at this moment (Moed 2005a, Van Raan 2006). Peer-review is costly and time consuming however, and in the UK, the Department of Health have moved towards a combination of peer-review and citation analysis to identify quality at the research funding stage (Van Leeuwen, Grant 2007, Nason, Grant et al. 2007), something mirrored in the latest changes to research funding for universities in the UK and Australia (Hodges 2006, Research Quality Framework 2006).

e. Esteem measures

As a long term measure of quality, there are several esteem metrics for researchers available. These include scientific awards, keynote speeches, invited lectures and journal editorship (Council for Medical Sciences 2002). These esteem metrics show off the quality of an individual over their career but do not link well to specific research findings or funding. There have been attempts to investigate researcher's CVs to identify if they are high quality researchers, but these essentially require peer-review of CVs and can be time consuming and costly.

Outreach

Since networking and outreach seems to be an important aspect of scientific impact (Wooding, Hanney et al. 2005), there are also indicators that investigate how knowledge is transferred in the academic community to identify how researchers work together to advance knowledge.. Co-author analysis of publications and patents allows identification of international and interdisciplinary collaboration (Glanzel, Schubert 2004). It is also possible to identify researchers spending time in other countries or laboratories (for instance on sabbatical) but this requires researchers to report this happening. To show how research is used by different academics, citation analyses can be modified to show the fields of research that are citing a publication. This gives an indication of the interdisciplinarity of the research by demonstrating the pickup of research outside the core discipline (Research Evaluation and Policy Project 2005).

Contextual and Structural

To understand the output of an institution, province or Canada as a whole, requires an analysis of the structure of research outputs. This includes the fields in which publications appear and where citations arise (identifying the applicability of research results across fields); the number of citers (giving an impression of the breadth of the field the publication is in); the activity index (which benchmarks publication volumes against the rest of the field, identifying how active an organization is in a field); and the position of journals that researchers publish in based on quality measures (usually the average citations per paper in the journals) (Research Evaluation and Policy Project 2005). It is also possible to address the structure of networks directly, which the Networks of Centres of Excellence already does for their members (R.A. Malatest and Associates Ltd., Circum Network Inc. 2007).

Capacity Building Indicators

Aside from producing new knowledge, health research funding also has a role in maintaining and improving the research capacity. At CIHR, this category includes the development and enhancement of research skills in individuals and teams (CIHR 2008). It is interesting to note that CIHR have toyed with the idea of rolling capacity building into advancing knowledge (Canadian Institutes for Health Research 2007), showing the close relationship between advancing knowledge and building research capacity.

Personnel

The most commonly associated item with capacity for research is human capacity. This includes data on PhD graduates in health research (Council for Medical Sciences 2002, King 2008), including trying to follow research graduates to identify where they use the skills gained through research (since this is

where impacts will occur) (Wolfson n.d.). It also includes the total numbers of research staff employed in health research (Statistics Canada 2008b), the numbers of researchers that are funded by particular funding organizations (e.g. (Birdsell, Asselbergs 2006)), the distribution of researchers (both in terms of geography and research field), the numbers of researching health professionals (since this is expected to improve the transfer of findings to practice) and the profile of researchers and graduates.

Capacity building also includes building receptor capacity in those organizations that need to use research and absorptive capacity in research organizations. Receptor capacity can be measured through annual surveying of research users to identify how they understand and use research (Denis, Lomas et al. 2008). Absorptive capacity represents the ability of those in research to take on board others' research findings and exploit that knowledge. It is commonly measured through collaborations (particularly industry - academia collaborations); R&D funding intensity; and co-authoring (Cockburn, Henderson 2003, Griffith, Redding et al. 2003, Schmidt 2005). It would also be possible to track the disciplines that researchers cite in publications as an example of absorptive capacity across disciplines (although this would not highlight research in the same discipline coming from other sources).

Funding

Although an input to research, funding can also be seen as a capacity building impact since prior research can attract additional research funds. For provinces this could be research funding from national organizations such as CIHR (Cockburn, Henderson 2003, Griffith, Redding et al. 2003, Schmidt 2005), or it could be matched funding agreements for any funder. It could also represent an increase in funding for a research field based on specific research (for example if HSFC can show that some of their funded research led to increases in heart disease funding from industry), this might also include specific actions such as CIHR opening a new institute based on that health problem. The clearest link for funding following research is through extension of specific research programs. It is often impossible to disaggregate the role of the individual within a group, particularly in relation to grant income (Council for Medical Sciences 2002, Birdsell, Asselbergs 2006), meaning that these measures tend to break down when trying to understand the impact of an individual on future funding.

Infrastructure

The third aspect of capacity building that can come through health research is through improved infrastructure. This could be in the form of new kit for research (such as MRIs or electron microscopes); through new or improved laboratories and buildings; or even through the creation of new databases and repositories for storing and sharing data. For infrastructure, there is currently information on the levels of funding for infrastructure that accompanies activity funding (e.g. through CFI), but there is little information on specific types of infrastructure such as new equipment, laboratories or databases for researchers. While data on funding from major infrastructure funders is available, information on infrastructure funding from other sources (e.g. University re-allocation of space, etc.) is more difficult to collect.

Informing Decision Making Indicators

In order for research to have an impact, it must change the way that a stakeholder acts. Networking is something that is increasingly being seen as important in improving research impacts (Gläser, Spurling et al. 2004) , making the way stakeholders are informed an important indicator to take into account. For CIHR, this category includes the impacts of research in the areas of science, public, clinical and managerial decision-making, practice and policy (Wooding, Hanney et al. 2005, Wooding, Hanney et al. 2004). This group of indicators are more difficult to identify, since they rely on more qualitative information about how people make decisions and what influences them. They also all suffer from the

problem that decisions are very rarely based on a single piece of research, and often have to entertain several inputs that are unrelated to research (contextual factors).

The first sub-category of informing decision making indicators is 'health related' decision making, which covers decision making in health care, public health, social care, other health related decision making (e.g. health and safety at work), and health related education (training of new health professionals and continuing medical education for health professionals). The second subcategory is research related decision making, which covers decisions about research funding allocations, research policies and researcher education. The third sub-category is health products industry decision making. The fourth sub-category is general public decision making which covers the decisions of advocacy groups such as patient groups and the way the public is educated about research.

Health related decision making

This includes the way that research informs health and social care practitioners. This can be identified through the guidelines (based on research findings) that inform practitioners (CIHR 2008). This method can also be used to investigate Cochrane Reviews and other publications that practitioners might use to get access to research findings. Other useful proxies would include analysis of professional bodies' newsletters to identify what new research is transferred to practitioners (although without citation of research results, this would have to be a peer-review process to identify the research); or whether practitioners are members of guideline committees (Grant, Cottrell et al. 2000). Alternatively, interactions that practitioners have with researchers (through meeting at conferences) and with other, research aware practitioners (learning from peers how to implement research findings) can be measured to establish contacts between research and health practice. This also applies to seminars for health professionals.

Continuing education for health practitioners often uses set curricula materials that cite research, and these can have citations analysed in much the same way as clinical guidelines are.

Health related decision making also includes the way healthcare managers and policy makers use research, mainly health services research, in their management decisions through indicators such as the number of commissioned systematic reviews to support policy making (Wooding 2008) or through surveying clinical managers on their awareness and use of research (Canadian Institutes for Health Research 2005). Other useful approaches to understanding how research informs institution managers include whether they are members of healthcare provider advisory groups, whether their hospital provides information on research use through the hospital accreditation process (Lavis, Ross et al. 2003), or whether there is evidence of research findings from HSR in hospital or healthcare institution performance measures.

At a higher level, there are healthcare decisions made by provincial or federal government policy makers and indicators for their use of research would include commissioning of research projects or systematic reviews of evidence to support policy making or the development of tax credits for R&D. Seminars for policy makers are one way to transfer knowledge from research to health related policy, whilst consultancy is another (this could be to national or international policy makers, such as Health Canada or the WHO). Alternately, secondments to policy making are another route for knowledge to transfer to policy makers. Some internet based measures have been proposed for identifying policy makers interaction with research; web hits by individuals with domain names suggesting a decision-maker organization and newsletter subscriptions from individuals with mailing addresses for decision-maker organizations have both been suggested (Accreditation Canada 2008), but these would require collecting data from research websites on usage which poses both collection and privacy issues.

An ideal way to identify the research use would be to perform a citation analysis of the evidence used in policy papers and there is the possibility that new bibliometric tools may be developed that would allow citation analysis of policy documents (Lavis, Ross et al. 2003) and Google Scholar may emerge as a useful tool for analysing citations in policy documents (Lewison 2004). However, this approach is complicated by a poor citation culture in policy papers and the lack of a database of policy papers. Another proxy for this would be the presence of researchers on government bodies (Bakkalbasi, Bauer et al. 2006, Noruzi 2005).

Research related decision making

This represents the way that research findings feed into scientific progress. At the level of an individual this could be represented by participation in decision making on future research programs for organizations – something that also applies to organizations to show that research has influenced their strategic directions. Individual academics may make decisions differently based on the strength of research findings, particularly in the case of responses to funding applications. Researchers involved in peer-review of funding and refereeing publications have a large role in informing the decisions on research direction taken by individuals (Research Evaluation and Policy Project 2005). Alternatively, it is possible to see how research informs research proposals by citation analysis of research proposals (although there is a danger of research proposals containing large proportions of self-citations and closely related researchers such as collaborators).

In universities, researcher participation in university management and research assessment is a useful indicator of research linking to institutional decision making. On a similar level, presence of researchers in the administrative functions of professional and learned societies can link research to the organizations processes (Research Evaluation and Policy Project 2005).

At the level of policies for government and health organizations, it is possible to identify what research is cited in their policy statements, but this may require significant data mining. For R&D taxation issues (commonly tax credits), policy papers rarely identify any research that informs the tax policy, so surveying of tax policy makers or case studies of policies would provide the only route to identifying any research impact on the policy (Research Evaluation and Policy Project 2005).

Education also uses research in teaching the next generation of health workers, researchers and those in professions that are required to use research findings. The indicators here are citations to publications in the textbooks or reading lists, and changes to teaching curriculums that identify the research that underpins a particular course or module.

Health products industry decision making

The **health products industry** has to make decisions based on research, since it generally requires a level of basic research to inform its own more clinical research. One proxy indicator of this is industry investment in academic research, since it shows the interest in a particular research area by industry. Alternatively, it is possible to collect information on the research that informs production stage reports in industry (Lavis, Ross et al. 2003, Lavis 2006). This kind of bibliometric analysis would identify the specific research that has informed the different aspects of industrial development of the research.

Counting licensed patents provides information on research that has been used by industry (Government of Canada 2008). This can be benchmarked against previous years or against internationally held patents (Treasury Board of Canada Secretariat 2007). There are already Canadian groups such as Science-Metrix who produce patent analyses for Canada (Organisation for Economic Co-Operation and Development 2007) and data on licensed patents is already maintained in Canada and reported on by Treasury Board {{401 Treasury Board of Canada Secretariat 2007}}.

Consultancy and transfer of researchers (including sandwich courses for research students) provides a route for research to enter industry. Identifying co-authorship of publications and patents also identifies where public research is informing the private sector (and vice-versa). Other proxies exist for identifying research collaboration including material transfer agreements (although these are seen in the research community as disincentives to collaborate since they slow down the collaborative research process) and co-location analyses to show where industry is located in relation to academic centres, a factor that facilitates translation of research between academia and industry (Science-Metrix 2008).

General public decision making is difficult to identify, since there are so many competing factors in decisions made by the general public. One way to address this is to use proxies for the public's actions such as media coverage of research, with indicators covering the scope and range of media coverage. This could be through TV and Radio appearances, or coverage in newspaper, magazine or website article as reported by researchers in some sort of survey – an approach being taken in the UK, splitting media appearances into international, national, provincial and local coverage, providing a system for establishing depth of impact (Gunasekara 2006). Ideally it would be possible to use this media coverage in a bibliometric analysis (Wooding 2008), but this is not currently a well-established field, therefore identifying media coverage remains something that researchers must report.

A second way is to try to identify the reactions of the public to specific public health campaigns through the final outcomes they arrive at (an example of this might be the increase in sales of condoms after public health campaigns about sexually transmitted diseases). It is also possible to survey the general public on how research has influenced their decision making, but this is often a difficult question to answer and proxies must be found to address the link of research to decision making.

Interactions for researchers with the public can primarily happen through public lectures, but can also include consultancy to advocacy and patient groups by researchers (Lewison 2004). For public lectures data could be collected through an expanded standard CV (Wooding 2008). It would also be useful to capture ad hoc lectures to schools or children by researchers (The Common CV System 2006).

For advocacy groups, research cited in publications (leaflets etc.) produced by advocacy groups, including patient organizations, can provide insight into the research results used by the organization in informing the public. Alternatively consultations to advocacy groups (formal or informal) can be captured through an expanded researcher CV (Research Evaluation and Policy Project 2005).

Health Indicators

In the CIHR categorization, this category encompasses advances in prevention, diagnosis, treatment and palliation when related to research (The Common CV System 2006), but the difficulty arises in how to relate findings to research. For example, CIHR identify changes in health as an indicator but then rely on 'special studies' to link those to research findings (CIHR 2008). As such, we have looked at the best ways to identify health and health systems improvements, and where possible have tried to link to research. Our categorisation, based on the ongoing work of the Canadian Institute for Health Information, splits into three major groupings: health status; determinants of health; and health system performance (Canadian Institutes for Health Research 2005).

Health status

The major drive behind health research is improving health, and this can be reflected mostly through the changes to the health status of individuals. This can be through reducing the number of people dying (reducing mortality) or making people healthier (reducing morbidity). In making people healthier, there can also be improved quality of life (as well as reduction in disease).

Morbidity

The most basic measures of morbidity are prevalence and incidence of the condition in question in the population (such as the cases/new cases per 1000 population). Prevalence and incidence are already collected for specific conditions by statistics Canada (Statistics Canada, Canadian Institute for Health Information 2008) and PHAC (including asthma; arthritis; blood pressure; cancer; chronic obstructive pulmonary disease; depression; diabetes; dementia; influenza; and stroke) and data on injury prevalence is collected by CIHI.

There are also instruments available to determine quality of life through functional state, including ability to complete Activities of Daily Living (ADLs) and Instrumental Activities of Daily Living (IADLs) (Statistics Canada 2008a); WHO forms SF-36, SF-8 on health status and the EuroQoL tool (Miller, Rejeski et al. 2000), which monitor health functions. Each of these can provide information on the functional state of an individual, but are less often used in evaluating research impacts on health.

For specific diseases and conditions, it is often easier to find an intermediate measure of improved health that can be easily assayed, rather than a direct change to the condition. There are several examples of intermediate health outcomes that can be easily measured and that relate to improved health. These include: lipid density measures that relate to heart disease; bone density measures that relate to osteoporosis (Brooks 1996); and birth-weight which can link to a number of future complications (Guidelines and Protocols Advisory Committee 2005). Linking intermediate measures to research can be simple for research that identifies new or improved measures, or linkages between a measure and a health condition. However, it is more difficult for research that improves the measure, such as public health research that leads to a reduction in LDL cholesterol, since the link between the research findings and the changed measure is difficult to make with many factors controlling the levels of cholesterol such as exercise, diet and genetics.

Mortality

Reducing mortality should be easy to measure, since at its most basic is counting the number of people dying from a condition. In Canada there is data on total mortality and that for specific conditions (e.g. cancer) or population groups (e.g. infants) available through Statistics Canada. Life expectancy is also a mortality based measure, and can be measured at different times of life (e.g. at birth or at 65 years). Building on life expectancy is the mortality measure used by CIHR to monitor health improvements in their evaluation framework, the Potential Years of Life Lost (PYLL) (Wilcox 2001). PYLL measures the years lost due to premature death (before age 75), and provides a measure of mortality than can be standardized across conditions (Canadian Institutes for Health Research 2005).

Quality adjusted mortality

As with providing quality of life measures to morbidity, this can be identifies for mortality. The most common measures are disability-adjusted life expectancy (DALE); disability-free life expectancy (DFLE); Health adjusted life expectancy (HALE); and the Health Utilities Index (HUI) which measures a number of quality of life, functional state and mortality indicators. Since DALYs and QALYs take into account anticipated years of life left as well as their quality, they are more commonly used in evaluations than the quality adjusted mortality measures mentioned here.

The Disability Adjusted Life Year (DALY) takes into account reductions in life span and reductions in functional state (disability) for specific conditions (Statistics Canada 2007). DALYs for conditions across nations are collected and compared by the WHO (World Health Organization 2008b). DALYs take into account the quality of life experienced by an individual through assessing their functional state (disability adjustment). There are other measures that also allow quality of life to be taken into account

(two-week disability days; disability/activity limitation; conditions causing activity limitation; health expectancy), but these are not as widely used as DALYs.

Along a similar theme to DALYs are Quality Adjusted Life Years (QALYs), which use a quality of life measure to determine changes in health due to particular health interventions (World Health Organization 2008a). Since QALYs link to interventions they can be more easily linked to research informing those interventions, something recently done in the UK (Phillips, Thompson n.d.). DALYs and QALYs both have the added benefit of being monetised on a regular basis, allowing a link to the economic impact of health improvement (to be discussed later). It is worth noting that QALYs can be used to assess the direct benefit of individuals involved in clinical trials (a small subset of health improvements but one with an automatic link to research).

Patient Reported Outcome Measures (PROMs) provide a patient's view based on a standardized questionnaire to determine patient views on quality of care and quality of life post-treatment. PROMs are used in the UK by the NHS as part of the outcome measures for improving healthcare (Buxton, Hanney et al. 2008).

Determinants of health

Public health in particular looks to directly affect the determinants of health, rather than the health condition itself, so it is pertinent to identify where health research can realistically modify determinants of health (since there will be determinants that health research cannot influence such as family history, agricultural production, ethnicity and socio-economic equity). CIHI and Statistics Canada collect a large amount of data on the numbers of people affected by specific determinants (Fitzpatrick, Bowling et al. 2006).

Non-modifiable risk factors

It seems inherently strange to talk of determinants that can be affected by health research and then to start with non-modifiable risk factors, but this leads to an important observation. Family history is the classic example of risk factors that can't be modified, but with the advent of genetic engineering and potential gene therapies, it may be that health research can have an impact on the genetic predisposition to disease in the future.

Modifiable risk factors

These include smoking, obesity, physical activity, alcohol consumption, diet etc. Each of these has metrics that exist for identifying how much of a risk it is to an individual. For smoking this could be initiation rates, consumption rates, or nicotine levels. For physical activity this could be hours exercise a week or metabolic levels of exercise; for obesity it is likely to be body mass index; for alcohol it may be consumption rates or binge drinking measures; for diet and nutrition it is most often self-reported data on consumption (although proxies would also exist at a population level through sales of healthy foods). CIHI collect much of this data for Canada already.

Social and cultural determinants

Those social determinants that can be modified (excluding issues such as ethnicity) are also open to health research influence. The two major areas identified are education, where there may be examples of new public health education campaigns in schools or changes to the education environment such as vending machine products; and social cohesion, where there may be improved social support networks or new community health workers and alliances (Statistics Canada, Canadian Institute for Health Information 2008).

Environmental determinants

Environment here is wider than simply the ambient environment (such as air or water pollution) and can include the built and work environments, family environments, and dangers within the environment such as second hand smoke or infectious disease vectors. Metrics that have been identified here previously include exposure levels for pollution or second hand smoke (Harter, Leier n.d.)(D'Amato, Liccardi et al. 2000)(Environment Canada 2005); teen pregnancy levels; prevalence of paths and cycle-paths; and worksite health promotion facilities.

Health system performance

The health system performance is also something that can be affected by health research, with research leading to reduced waiting times (accessibility of care); improved patient satisfaction (acceptability of care); improved effectiveness of treatments (e.g. adherence to clinical guidelines); improved efficiency of treatments (e.g. inflow-outflow rates for patients); and improved safety (e.g. reduced medical errors). Again, it is worth re-iterating that the majority of these indicators show the health or health system improvements (or otherwise), but do not relate them to research. CIHI split the performance of the health system into eight groups (Statistics Canada, Canadian Institute for Health Information 2008), covered below.

Acceptability

CIHI define this as the health system being acceptable to patients (Canadian Institute for Health Information 2008a), and this is most easily identified through surveys of patient satisfaction (Canadian Institute for Health Information 1999). Obviously, as with any self-reported data there are inherent problems with this metric in it susceptibility to reporting bias, something experienced in many health self-reporting surveys (Niagara Health System n.d.). Linking this acceptability to research can only be done through case studies of what is more acceptable and why.

Accessibility

This is considered to be the need for the health system to provide for anyone at the right place, the right time on based on need (Steinbrook 2006, Leroux, Rizzo et al. 2003). As such this brings in issues of equity of treatment and the ethical requirements upon a health system. Timeliness of treatment is often covered through waiting times analyses (Canadian Institute for Health Information 1999) and has formed part of Statistics Canada's monitoring of access to the health system (The Canadian Press 2007) and are also collected already by CIHI (Sanmartin, Gendron et al. 2004). This report covers accessing physicians and routine care, as well as specialist secondary care.

Using appointment statistics could provide additional information on the public's use of health services (since appointment data exists at all levels of healthcare including dentistry and specialisms such as physiotherapy). Appointment statistics would require collecting data from primary care providers on the time to get appointments. This could be done through the Access Response Index (AROS), which counts the number of days until the next available routine appointment, with any clinician, once during every normal working day (Canadian Institute for Health Information 2008b).

Since telephone health advisory services also exist, the way the public accesses those should also be collected (perhaps through random selection surveying after the content section of the call is finished as is often seen in the private sector) (Jones, Elwyn et al. 2003).

Equity of access is vitally important in a representative health system, and can be a difficult issue to address, but Health Canada attempted to do so in 2001 (Marklund, Bengtsson et al. 1990). This broke down access into specific groups and compared their access. This is a good way to address the issue, but to be used as a metric for access it would have to use health service indicators that are well understood

in different groups, such as revascularization or statin use). By utilizing well understood health services or interventions we may also find it easier to link equity to research findings. Of course, specific HSR projects that investigate equity and access can be directly evaluated to show how HSR research has influenced access using the access statistics for specific groups.

Appropriateness

The appropriateness of the system can be defined as the system providing the highest quality interventions based on the best evidence (i.e. giving the most appropriate treatment). This can be estimated through identifying adherence to clinical guidelines for treatments. This has the added benefit that it provides a direct link to the research supporting the guidelines and how it makes the health system more appropriate through improved practice. Clinical audits can provide information on how well services conform to guidelines (Health Canada 2001). CIHI uses caesarean section births as a measure of appropriateness (Godwin 2001), but this is a measure that does not take into account when c-sections are necessary, only working on the assumption that a reduction to zero is desirable (clinically unlikely).

Competence

Competence is considered the application of skills appropriately in the health system (Statistics Canada, Canadian Institute for Health Information 2008) and can be approximated by identifying when the system is not appropriate, for example through counts of civil law suits against the health system.

Continuity

Continuity of care, defined as the 'extent to which health care services over time are perceived as a coherent and connected succession of events consistent with a patient's medical needs and personal context' (Canadian Institute for Health Information 1999), is a desired goal of the health system (Centre for Health Services and Policy Research 2004) and one that is difficult to measure. In general it is associated with primary care guiding the patient through the health care system, and it is likely that once electronic health records are better established, anonymized data from those could help to identify whether patients have had their care follow a continuous path. Currently however, the only measures of continuity are self-reported survey data on satisfaction with the care pathway and administrative records from health professionals. Surveying patients to identify their perception of the continuity of their care provides a method for identifying experience for patient groups (Health Canada 2005). Using administrative data; data collection is simple but the results often do not identify well with experience of continuity, the desirable aspect of continuity (Centre for Health Services and Policy Research 2004).

Effectiveness

The effectiveness of a health system comes from its ability to provide the best healthcare, achieving the desired result (Centre for Health Services and Policy Research 2004). CIHI use 30 day-in hospital mortality as their main indicator for effectiveness (Canadian Institute for Health Information 1999); understandable since it determines the effectiveness of hospitals in preventing death. This measure can also be applied to specific conditions so is useful for a number of conditions or interventions. Where there are guidelines that are followed, leading to a more effective system for conditions we can trace back to the research underpinning the guidelines. As mortality is not going to be an issue for all conditions, CIHI also identify re-admission rates for conditions as a measure of effectiveness (Statistics Canada, Canadian Institute for Health Information 2008). Both of these metrics suffer from a focus on secondary health care, and at present there are no good indicators identified for the primary care sector.

Efficiency

Defined by CIHI as 'achieving the desired result with the most cost-effective use of resources' (Statistics Canada, Canadian Institute for Health Information 2008) p.3), it is interesting to note that the measures that CIHI identify for efficiency do not refer to the use of resources. For example, they include hospitalization rate for ambulatory care sensitive conditions (Canadian Institute for Health Information 1999) since this identifies conditions that can often be dealt with in a primary or social care setting. However, without understanding whether the hospitalizations are required, it is an assumption that reducing numbers is the most efficient way of dealing with ambulatory conditions.

A more appropriate approach may be to look at the actual vs. expected length of in-patient stay, since this identifies a reduction in resource use based on expert opinion prior to any treatment. Also, HTA studies would be useful in identifying the efficiency of the health system, since they identify in monetary terms the effectiveness of specific treatments for given conditions. Also, HTA as a research discipline provides a clear link to research for efficiency measures of health services (at least those that are recommended)(Statistics Canada, Canadian Institute for Health Information 2008).

Alternatively, collecting data on the inputs to healthcare services and on the different factors identified as outputs (e.g. available beds, emergency admissions etc.) provides the information that can be fed into a model to identify efficiency, such as a Stochastic Frontier Analysis (SFA) (Jacob, McGregor 1997), an approach used in Ireland (McGlynn, Shekelle et al. 2008). Much of this data is already collected for healthcare providers, such as the costs of healthcare provision and the different outputs of health care.

Safety

Safety is considered to be reducing the risks of an intervention or healthcare environment (Lordan 2007). Within health system competence we are already identifying when serious errors occur and safety is compromised, but there are also specific measures of error that can be used to measure safety: the number adverse drug effects, adverse surgical effects or reactions to anaesthesia; and hospital acquired infections (HAIs) are particularly topical examples. The Canadian Patient Safety Initiative (CPSI) performs audits of providers to identify numbers of adverse effects in general (Canadian Institute for Health Information 2008a, Canadian Institute for Health Information 1999), although this only covers a small proportion of providers and is resource intensive since it requires identification of all adverse effects using expert opinion.

Economic Indicators

These have already been identified in the section on the payback model for evaluating health research impacts above. Within each of these four types of economic impact are a number of indicators that can help identify research's role in economic benefits.

As mentioned, HTA studies can be a good indicator in determining the research link to this clearly only goes to one level of attribution, showing what research has been involved in the cost savings due to the technology development, but does not attribute specific levels of credit to any particular research. Recent work in the UK (Baker, Norton et al. 2004) has linked research to economic impacts through monetisation of QALY gains (the chosen health impact).

Alternatively, it is possible to take a human capital approach and perform economic analyses to identify the economic benefits from research, but this has pitfalls in that it ignores employees being replaced by unemployed people or through migration and it limits benefits from improved health to those of working age (Buxton, Hanney et al. 2008).

Benefits to the economy from commercial development has been the economic benefit area where most attention has been placed in terms of indicators. This is because there are obvious areas to collect information that link to research, such as patent licensing, spin-off companies, drug sales and health industry employment, all of which are collected by Statistics Canada.

The **broader benefits to society of the health gain from medical advances** are the most problematic of the economic benefits to pin down with indicators because they rely on so many assumptions. However, the economic valuing of health benefits in the form of DALYs and QALYs is now a well established practice (Nason, Janta et al. 2008), and even the value of a statistical life used in the Exceptional Returns studies has been the subject of meta-analyses (Dolan, Shaw et al. 2004). The problems arise predominantly because although there is extensive work on these types of economic values for life and health, there are still disagreements over the way the valuations are performed (Bellavance, Dionne et al. 2007).

Broad Economic and Social Benefits

CIHR focus on the economic benefits here and divide this category into three subcategories: commercialization of discoveries; direct cost savings; and human capital gains (Buxton, Hanney et al. 2004). These are based on four major areas of economic impact identified in a review of economic impacts from health research for the WHO (CIHR 2008). Direct cost savings to health care systems can be identified using the cost savings due to a new technology that can then be linked to research that underpins the technology (Buxton, Hanney et al. 2004). Benefits to the economy from a healthy work force can be identified through proxies such as the number of sick days in different professions (collected by Statistics Canada). 'Benefits to the economy from commercial development' has been the area where most attention has been placed because there are obvious indicators that link to research, such as patent licensing, spin-off companies, drug sales and health industry employment, all of which are collected by Statistics Canada. The broader benefits to society of the health gain from medical advances have used monetisation of DALYs and QALYs (Jacob, McGregor 1997)(Dolan, Shaw et al. 2004), but these are the problematic because they rely on so many assumptions leading to disagreements over the way the valuations are performed (Bellavance, Dionne et al. 2007).

We have identified three sub-categories for the economic impacts of health research: the benefit accrued through the action of research rather than the outputs of research; commercialization of research results; and the net benefit of improving health. We also consider the impacts of research on well being and the social impacts of research.

Activity of research

The first and most obvious example of an activity benefit is the employment of researchers. Other ways activity can provide benefit include an effect on the health of research participants (captured in health benefits), the employment of and retention of clinical staff (capacity building), the funding brought into the country or province from international firms (capacity building), and the ability to take on board new knowledge from other studies ('absorptive capacity').

To identify the economic benefit of employment, what has often been used has been counts of those employed in health research, but to identify a benefit, this needs to take into account what people would be doing were they not involved in research. Economic rent can take into account the differential impacts of human resources versus other industries, i.e. the benefit of employing people in biomedical research as opposed to employing them in next best use of their skills (Buxton, Hanney et al. 2004). This differential measure is the concept of labour rents – the excess earnings above the marginal cost of the labour. For example if a researcher is paid $30,000 and the next best job they could get would pay $25,000, then the economic rent is $5,000 p/a. (Garau, Sussex 2007, Martin, Tang June 2007). As an

indicator, economic rent can be applied as long as baseline data on research inputs can be identified through Statistics Canada data ((Garau, Sussex 2007), and the outputs data required could be captured through Rx&D (Science, Innovation and Electronic Information Division 2008). This approach has been used in recent work in the UK to show how medical research provides economic value to the economy (Rx&D 2007, Rx&D 2006b) and has been used for the pharmaceutical industry (Buxton, Hanney et al. 2008) so is applicable to both public and private sector employment in research. This approach is not a simple to use metric however and requires good data on investments made into health research as well as expertise to undertake the study. Because economic rent involves a number of assumptions and data collection issues, it is sensitive to poor data if there are small numbers involved.

Commercialization

Previous evaluations have taken a bottom-up approach to identifying the economic benefits from commercialization, using measures such as patent licences, research product sales revenues and spin-out companies (Garau, Sussex 2007). Summing the dollars spent on licensing patents held by Canadian organizations/individuals allows us to identify the economic impact of licensed patents and to link this impact to specific research findings (Muir, Arthur et al. 2005).Sales revenues of products developed in Canada provide a simple measure of the economic impact of health products (Science-Metrix 2008, Byrd 2002). However there are difficulties in linking sales revenues to research findings due to the other factors that affect sales. Using the valuation of portfolios of new spin out companies and the sales of spin outs to provide the value to the economy of spin-outs at any given point (annually) could give an indication of the economic value of new companies coming out of research. The number and nature of spin out companies can be relatively easily identified (Science-Metrix 2008, Byrd 2002) but accessing valuation of new spin outs may be difficult, although presumably could be made available through venture capital firms that support the spin out companies. Any use of this indicator over a number of years would need to take into account market conditions that could change the value of new companies drastically, but economic techniques can deal with these changes to conditions.

This method provides collectable data, but misses out the other aspects of commercialization that occur because of research funding, for example through publicly funded researchers interaction with industry in a consultancy role (informing decision making). To try and understand the less measurable impacts of research on commercialization it is necessary to look at the overall behaviour of the health products industry and its relationship with R&D. This can be done using the economic rent approach but focusing on the producer rents (the economic benefits on top of expected revenues) and spillover effects (knock-on affects improving economic returns outside of the specific R&D being undertaken) from R&D commercialization (Lonmo 2008). Data on producer rents can be identified through revenue statements and expected levels of profit for organizations. Spillover effects can be identified through an analysis similar to that used in the UK (Garau, Sussex 2007) in which the researchers identified public and private investments in R&D (since there are complementarities between public and private funding for R&D; (Buxton, Hanney et al. 2008)) and then estimated the private and social rates of return on the R&D based on previous studies of rates of return when changes occur in public and private R&D inputs (Congressional Budget Office 2006). Using this method captures the total effects that changes to R&D funding has on commercial profits rather than just identifying the benefits of patents, sales and spin-outs.

Net health benefit

Since it is important to understand not only the benefits of improved health, but also how those relate to the costs of improving health, we suggest using the net economic benefit as an approach to understanding health gain benefits. This essentially uses the costs of implementing the health improvement arrived at through research from the health gain value to give a measure of health

improvement per dollar (Buxton, Hanney et al. 2008). If measured using QALYs or DALYs, then this value can be a $ value, since QALYs and DALYs can be monetised (albeit using a controversial methodology – e.g. (Buxton, Hanney et al. 2008)). This would mean that the net health benefit could be compared to the $ payback on other uses of funding aside from health-related funding. QALYs provide the benefit of a linkage to specific interventions, making them more suited to a linkage to research (Bingham 2001). However, DALYs provide monetarised information on conditions rather than treatments and are a more ubiquitous measure of health improvement (and are also simple to calculate for situations where data does not currently exist), meaning DALYs could provide a better overview of health improvement despite disputes over the ways to monetise DALYs (Buxton, Hanney et al. 2008). As with the use of QALYs and DALYs, improvements in health measured through PROMs gained could be divided by the cost of achieving that health gain. This approach suffers from two drawbacks. First, that PROMS have not been used in this kind of approach before (meaning that there is no data to compare to); and second, that PROMs have not been monetised so this measure can only be compared to other PROMs measures, not any funding for non-health related activities. Identifying the cost of implementing the health gain has been performed using specific evidence based interventions that have altered health since work on identifying monetised net health gain has focused on medical research (Buxton, Hanney et al. 2004). This is a complex methodology that is again reliant on getting good data on health improvements and implementation costs as well as requiring expert analysis, implying that this method requires resource allocation to perform.

Well being indicators

Well being is a difficult concept to measure and here we identify two comprehensive measure and two aspects of well being to highlight the importance of measuring well being as an impact of health research.

Human Resources and Social Development Canada (HSRDC) have multiple indicators of well-being, brought together in an annual report of well being for Canada, although these currently have no links to research (health or otherwise) except through the 'health' section of the well-being indicators (these indicators are covered in the Health Impacts category). Since this data is already collected and publicly accessible, this is an attractive starting point to access information on changes in well being in Canada. The Canadian Index of Well-being (CIW) is a new index with multiple parameters that aims to identify a single value for well-being in Canada – aiming to be the well-being equivalent of GDP (Buxton, Hanney et al. 2008). Without having the index in place, it is difficult to assess how useful it will be in linking to Canadian health research, but it could provide useful additional data on well being outcomes in Canada.

As individual aspects of well-being there are a number of areas that data could be collected. They include happiness and levels of social isolation. There are a number of measures used to assess happiness such as the self-report happiness scales used by Statistics Canada and the short depression-happiness scale (The Atkinson Charitable Foundation 2005). There is a clear difficulty here in linking happiness changes to health research, but without collected data on happiness this link will never be possible. By using loneliness scales for measuring social isolation of individuals such as the UCLA loneliness scale (Joseph, Linley et al. 2004) we can access information on the social isolation of individuals. Again, linking changes in social isolation, as with happiness, to health research is currently very difficult.

Social benefit indicators

It is important to keep in mind that research can have social impacts as well as economic ones as final outcomes. Social benefits that arise from health research can be incredibly wide ranging, from changes to the built environment (more playing fields for schools) through to improved health education for the public (improving educational attainment and understanding of science). Since the number of potential

social indicators is so high, we provide an example of an indicator, rather than indicators for all possible social impacts.

The causality of socio-economic status to health outcomes is well known (McWhirter 1990) but it is not understood if health research can alter socio-economic status. Identifying socio-economic status of individuals in Canada should be collected to identify if changes in socio-economic status correlate with research impacts, something that would have to be assessed through specific research studies.

References

ACCREDITATION CANADA, 2008. *2008 Canadian Health Accreditation Report: Based on analysis from surveys conducted in 2007.* Ottawa, Ontario: Accreditation Canada.

BAKER, G.R., NORTON, P.G., FLINTOFT, V., BLAIS, R., BROWN, A., COX, J., ETCHELLS, E., GHALI, W.A., HEBERT, P., MAJUMDAR, S.R., O'BEIRNE, M., PALACIOS-DERFLINGHER, L., REID, R.J., SHEPS, S. and TAMBLYN, R., 2004. The Canadian Adverse Events Study: the incidence of adverse events among hospital patients in Canada. *Canadian Medical Association Journal,* **170**(11), pp. 1678-1686.

BAKKALBASI, N., BAUER, K., GLOVER, J. and WANG, L., 2006. Three options for citation tracking: Google Scholar, Scopus and Web of Science. *Biomedical Digital Libraries,* **3**(7),.

BELLAVANCE, F., DIONNE, G. and LEBEAU, M., 2007. *The Value of a Statistical Life: A Meta-Analysis with a Mixed Effects Regression Model.* Working paper 06-12. Montreal, Quebec: Canada Research Chair in Risk Management, HEC.

BINGHAM, T.H., 2001. Estimating Economic Benefits from ATP Funding of New Medical Technologies. In: C.W. WESSNER, ed, *The Advanced Technology Program.* Washington, D.C.: National Academy Press, pp. 211-222.

BIRDSELL, J.M. and ASSELBERGS, M., 2006. *Health Research Funding In Manitoba: A Description and Comparison Across Canada.* Manitoba, Canada: On Management Health Group.

BORNMANN, L., MUTZ, R. and DANIEL, H., 2008. Are There Better Indices for Evaluation Purposes than the *h* Index? A Comparison of Nine Different Variants of the *h* Index Using Data from Biomedicine. *Journal of the American Society for Information Science and Technology,* **59**(5), pp. 830-837.

BROOKS, R., 1996. EuroQol: the current state of play. *Health Policy,* **37**(1), pp. 53-72.

BUTLER, L., 2002. A list of published papers is no measure of value. *Nature,* **419**(6910), pp. 877.

BUXTON, M., HANNEY, S. and JONES, T., 2004. Estimating the economic value to societies of the impact of health research: a critical review. *Bull World Health Organ,* **82**(10), pp. 733-739.

BUXTON, M., HANNEY, S., MORRIS, S., SUNDMACHER, L., MESTRE-FERRANDIZ, J., GARAU, M., SUSSEX, J., GRANT, J., WOODING, S., NASON, E., ISMAIL, S. and KAPUR, S., 2008. Medical Research – What's it worth? Estimating the economic benefits from medical research in the UK. Report to the UK Evaluation Forum edn. London, UK: .

BYRD, C.A., 2002. *Profile of Spin-off Firms in the Biotechnology Sector: Results from the Biotechnology Use and Development Survey - 1999.* 88F0006XIE No. 04. Ottawa, Canada: Statistics Canada.

CANADIAN INSTITUTE FOR HEALTH INFORMATION, 2008a. *Health Indicators 2008: Definitions, Data Sources and Rationale.* Ottawa, Ontario: CIHI.

CANADIAN INSTITUTE FOR HEALTH INFORMATION, 2008b. *Wait Time Tables—A Comparison by Province, 2008.* Ottawa, Ontario: CIHI.

CANADIAN INSTITUTE FOR HEALTH INFORMATION, 1999. *National Consensus Conference on Population Health Indicators.* Ottowa, Ontario: CIHI.

CANADIAN INSTITUTES FOR HEALTH RESEARCH, November 2007, 2007-last update, CIHR framework to measure the impact of health research: "the socio-economic returns of medical research workshop" [Homepage of Swedish Research Council], [Online]. Available: http://www.vr.se/download/18.5ec4ab911161ebf14dc80005426/CIHR+Borbey.pdf [July 22nd, 2008].

CANADIAN INSTITUTES FOR HEALTH RESEARCH, 2005. *Developing a CIHR Framework to Measure the Impact of Health Research.* Ottawa, Ontario: CIHR.

CANADIAN INSTITUTES OF HEALTH RESEARCH, September 17th 2007, 2007-last update, an overview of CIHR [Homepage of Camadian Institutes of Health Research], [Online]. Available: http://www.cihr-irsc.gc.ca/e/30240.html#slide1_e [June 26th, 2008].

CENTRE FOR HEALTH SERVICES AND POLICY RESEARCH, 2004. *Big steps, small leaps: Defining and measuring continuity of care.* CHSPR 04:23S. Vancouver, B.C.: CHSPR.

CIHR, 2008. Payback categories. Ottawa, Canada: CIHR.

COCKBURN, I.M. and HENDERSON, R.M., 2003. Absorptive Capacity, Coauthoring Behavior, and the Organization of Research in Drug Discovery. *The Journal of Industrial Economics,* **46**(2), pp. 157-182.

CONGRESSIONAL BUDGET OFFICE, 2006. *Research and Development in the Pharmaceutical Industry.* Washington, D.C.: Congressional Budget Office.

COUNCIL FOR MEDICAL SCIENCES, 2002. *The societal impact of applied health research: Towards a quality assessment system.* Amsterdam, the Netherlands: Royal Netherlands Academy of Arts and Sciences.

D'AMATO, G., LICCARDI, G. and D'AMATO, M., 2000. Environmental risk factors (outdoor air pollution and climatic changes) and increased trend of respiratory allergy. *Journal of investigational allergology & clinical immunology,* **10**(3), pp. 123-128.

DENIS, J., LOMAS, J. and STIPICH, N., 2008.
Creating receptor capacity for research in the health system: the Executive Training for Research Application (EXTRA) program in Canada. *Journal of Health Services Research Policy,* **13**(Supplement 1), pp. 1-7.

DOLAN, P., SHAW, R., TSUCHIYA, A. and WILLIAMS, A., 2004. QALY maximisation and people's preferences: a methodological review of the literature. *Health Economics,* **14**(2), pp. 197-208.

ENVIRONMENT CANADA, November 4th, 2005-last update, urban air quality [Homepage of Environment Canada], [Online]. Available: http://www.ec.gc.ca/soer-ree/English/Indicator_series/new_issues.cfm?tech_id=31&issue_id=8 [August 30th, 2008].

ESSENTIAL SCIENCE INDICATORS, 2008-last update, new hot papers [Homepage of Thomson], [Online]. Available: http://esi-topics.com/nhp/index.html [August 20th, 2008].

FITZPATRICK, R., BOWLING, A., GIBBONS, E., HAYWOOD, K., JENKINSON, C., MACKINTOSH, A. and PETERS, M., 2006. *Health Outcome Indicators: A structured review of patient-reported measures in relation to selected chronic conditions, perceptions of quality of care and carer impact.* Oxford, UK: Patient-reported Health Instruments Group; National Centre for Health Outcomes Development.

GARAU, M. and SUSSEX, J., 2007. *Estimating Pharmaceutical Companies' Value to the National Economy: Case Study of the British Pharma Group.* London, UK: Office of Health Economics.

GLANZEL, W. and SCHUBERT, A., 2004. Analyzing Scientific Networks Through Co-authorship. In: H.F. MOED, W. GLÄNZEL and U. SCHMOCH, eds, *Handbook of Quantitative Science and Technology Research. The Use of Publication and Patent Statistics in Studies of S&T Systems.* Dordrecht (the Netherlands: Kluwer Academic Publishers, pp. 257-276.

GLÄSER, J., SPURLING, T.H. and BUTLER, L., 2004. Intraorganisational evaluation: are there 'least evaluable units'? *Research Evaluation,* **13**(1), pp. 19-32.

GODWIN, M., 2001. Conducting a clinical practice audit: Fourteen steps to better patient care. *Canadian Family Physician,* **47**, pp. 2331-2333.

GOVERNMENT OF CANADA, May 9th, 2008-last update, food and consumer product safety action plan: health products [Homepage of Government of Canada], [Online]. Available: http://www.healthycanadians.ca/pr-rp/dpaper-papier3_e.html [August 21st, 2008].

GRANT, J., COTTRELL, R., CLUZEAU, F. and FAWCETT, G., 2000. Evaluating "payback" on biomedical research from papers cited in clinical guidelines: applied bibliometric study. *British Medical Journal,* **320**, pp. 1107-1111.

GRIFFITH, R., REDDING, S. and VAN REENEN, J., 2003. R&D and Absorptive Capacity: Theory and Empirical Evidence. *Scandinavian Journal of Economics,* **105**(1), pp. 99-118.

GUIDELINES AND PROTOCOLS ADVISORY COMMITTEE, 2005. *Bone Density Measurement in Women.* Victoria, BC: British Columbia Ministry of Health.

GUNASEKARA, C., 2006. Academia and industry: The generative and developmental roles of universities in regional innovation systems. *Science and Public Policy,* **33**(2), pp. 137-150.

HARTER, J. and LEIER, M., n.d. *No Time To Care: Community Health Workers and the New Economy.* Vancouver, B.C.: Simon Fraser University.

HEALTH CANADA, 2005. *Canada's Health Care System.* Ottawa, Ontario: Health Canada.

HEALTH CANADA, 2001. *"Certain Circumstances": Issues in Equity and Responsiveness in Access to Health Care in Canada.* Ottawa, Ontario: Health Canada.

HEALTH ECONOMICS RESEARCH GROUP, 2008-last update, HERG reports: HERG research reports [Homepage of Brunel University], [Online]. Available: http://www.brunel.ac.uk/about/acad/herg/publications/hergreports [AUgust 24th, 2008].

HODGES, I., 2006. The RAE is dead - long live metrics. *The Independent on Sunday,* **Higher Education**, pp. http://www.independent.co.uk/news/education/higher/the-rae-is-dead--long-live-metrics-473876.html.

INSTITUTE FOR SCIENTIFIC INFORMATION, n.d.-last update, the thomson scientific impact factor [Homepage of Thomson Scientific], [Online]. Available: http://scientific.thomsonreuters.com/free/essays/journalcitationreports/impactfactor/ [August 24th, 2008].

JACOB, R. and MCGREGOR, M., 1997. Assessing the impact of health technology assessment. *International Journal of Technology Assessment in Health Care,* **13**(1), pp. 68-80.

JONES, W., ELWYN, G., EDWARDS, P., EDWARDS, A., EMMERSON, M. and HIBBS, R., 2003. Measuring access to primary care appointments: a review of methods. *BMC Family Practice,* **4**(8),.

JOSEPH, S., LINLEY, P.A., HARWOOD, J., LEWIS, C.A. and MCCOLLAM, P., 2004. Rapid assessment of well-being: The Short Depression-Happiness Scale (SDHS). *Psychology and Psychotherapy: Theory, Research and Practice,* **77**, pp. 463-478.

KELLEY, E. and HURST, J., 2006. *Health Care Quality Indicators Project Conceptual Framework Paper.* OECD Health Working Papers No. 23. Paris, France: OECD.

KING, D., 2008. *Doctoral Graduates in Canada: Findings from the Survey of Earned Doctorates, 2004/2005.* No. 065. Ottawa, Ontario: Statistics Canada.

LAVIS, J., ROSS, S., MCLEOD, C. and GILDINER, A., 2003. Measuring the impact of health research. *Journal of health services research & policy,* **8**(3), pp. 165-170.

LAVIS, J.N., 2006. Research, public policymaking, and knowledge-translation processes: Canadian efforts to build bridges. *The Journal of continuing education in the health professions,* **26**(1), pp. 37-45.

LEROUX, J., RIZZO, J. and SICKLES, R., 2003. *Health, Mental Health and Labor Productivity: The Role of Self-Reporting Bias.* US: Rice University/Ohio State University.

LEWISON, G., 2004. Citations to Papers from Other Documents: Evaluation of the Practical Effects of Biomedical Research. In: H.F. MOED, W. GLANZEL and U. SCHMOCH, eds, *Handbook of Quantitative Science and Technology Research.* Dordrecht, Netherlands: Kluwer Academic Publishers, pp. 457-472.

LONMO, C., 2008. Biotechnology spinoffs: transferring knowledge from universities and government labs to the marketplace. *Innovation Analysis Bulletin, Statistics Canada,* **10**(1),.

LORDAN, G., 2007. *Measuring efficiency in health care: an application to out of hours primary care services in the island of Ireland.* HEDG Working Paper 07/08. York, UK: Health Economics and Data Group, University of York.

MARKLUND, B., BENGTSSON, C., BLOMKVIST, S., FURUNES, B., GACKE-HERBST, R., SILFVERHIELM, B. and WENNBERG, G., 1990. Evaluation of the Telephone Advisory Activity at Swedish Primary Health Care Centres. *Family Practice,* **7**(3), pp. 184-189.

MARTIN, B.R. and TANG, P., June 2007. *The benefits from publicly funded research.* Paper No. 161. Science and Technology Policy Research, University of Sussex.

MCGLYNN, E.A., SHEKELLE, P.G., CHEN, S., GOLDMAN, D., ROMLEY, J., HUSSEY, P., DE VRIES, H., WANG, M., TIMMER, M., CARTER, J., TRINGALE, C. and SHANMAN, R., 2008. *Identifying, Categorizing, and Evaluating Health Care Efficiency Measures.* Rockville, Maryland: Agency for Healthcare Research and Quality, USDHHS.

MCWHIRTER, B.T., 1990. Factor analysis of the revised UCLA loneliness scale. *Current Psychology,* **9**(1), pp. 56-68.

MILLER, M.E., REJESKI, W.J., REBOUSSIN, B.A., TEN HAVE, T.R. and ETTINGER, W.H., 2000. Physical activity, functional limitations, and disability in older adults. *Journal of the American Geriatrics Society,* **48**(10), pp. 1264-1272.

MOED, H.F., 2005a. Citation Analysis in Research Evaluation. Dordrecht, Netherlands: Springer.

MOED, H.F., 2005b. Statistical relationships between downloads and citations at the level of individual documents within a single journal. *Journal of the American Society for Information Science and Technology,* **56**(10), pp. 1088-1097.

MOHER, D. and BERNSTEIN, A., 2004. Registering CIHR-funded randomized controlled trials: a global public good. *Canadian Medical Association Journal,* **171**(7), pp. 750-751.

MUIR, R., ARTHUR, E., BERMAN, T., SEDGLEY, S., HERLICK, Z. and FULLGRABE, H., 2005. *Metrics for Research Commercialisation: A Report to the Coordination Committee on Science and Technology.* Canberra, Australia: Coordination Committee on Science and Technology.

NASON, E., GRANT, J. and VAN LEEUWEN, T.N., 2007. *Bibliometric analysis of highly cited publications of health research in England, 1997-2003:Theme specific HCPs in England.* WR-509-DH. Cambridge, UK: RAND Europe.

NASON, E., JANTA, B., HASTINGS, G., HANNEY, S., O'DRISCOLL, M. and WOODING, S., 2008. *Health Research – Making an Impact: The Economic and Social Benefits of HRB Funded Research.* Dublin: Ireland: HRB Ireland.

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, February 13th, 2008-last update, NIST scientific and technical databases [Homepage of NIST], [Online]. Available: http://www.nist.gov/srd/ [October 31st, 2008].

NCBI, n.d. 2008a-last update, entrez cross-database search page [Homepage of NCBI], [Online]. Available: http://www.ncbi.nlm.nih.gov/sites/gquery [October 31st, 2008].

NCBI, April 2nd, 2008b-last update, GenBank overview [Homepage of NCBI], [Online]. Available: http://www.ncbi.nlm.nih.gov/Genbank/ [October 31st, 2008].

NIAGARA HEALTH SYSTEM, n.d. n.d.-last update, qaulity initiatives: patient satisfaction [Homepage of Niagara Health System], [Online]. Available: http://www.niagarahealth.on.ca/quality_safety/quality/patient_satisfaction.html [August 28th, 2008].

NORUZI, A., 2005. Google Scholar: The New Generation of Citation Indexes. *Libri,* **55**(4), pp. 170-180.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, 2007. *OECD Science, Technology and Industry Scoreboard 2007: Innovation and Performance in the Global Economy.* Paris, France: OECD.

OSTRY, A., 1999. *Explaining the Link Between Socioeconomic Status and Health.* HPRU 99:12D. Vancouver, B.C.: Centre for Health Services and Policy Research.

PHILLIPS, C. and THOMPSON, G., n.d. *What is a QALY?* Volume 6, Number 1. Newmarket, UK: Hayward Medical Communications.

R.A. MALATEST AND ASSOCIATES LTD. and CIRCUM NETWORK INC., 2007. *Evaluation of the Networks of Centres of Excellence Program: Evaluation report.* Ottawa, Ontario: NCE Steering Committee.

RESEARCH EVALUATION AND POLICY PROJECT, August 2nd, 2008-last update, about REPP [Homepage of Australian National University], [Online]. Available: http://repp.anu.edu.au/ [July 31st, 2008].

RESEARCH EVALUATION AND POLICY PROJECT, 2005. *Quantitative Indicators for Research Assessment - A Literature Review.* Review No. 05/1. Canberra, ACT: Research School of Social Sciences, Australian National University.

RESEARCH QUALITY FRAMEWORK, 2006. Assessing the quality and impact of research in Australia. The Recommended RQF.

RX&D, 2007. *The Pharmaceutical Industry in Canada.* September. Ottawa, Ontario: Rx&D.

RX&D, May 4th, 2006a-last update, clinical trials portal [Homepage of Rx&D], [Online]. Available: http://www.canadapharma.org/Pharm_comm/Trials/index_e.html [August 21st, 2008].

RX&D, January 23rd, 2006b-last update, did you know? our community [Homepage of Rx&D], [Online]. Available: http://www.canadapharma.org/DidYouKnow/index_e.html [July 27th, 2008].

SANMARTIN, C., GENDRON, F., BERTHELOT, J.M. and MURPHY, K., 2004. *Access to Health Care Services in Canada, 2003.* Ottawa, Ontario: Statistics Canada.

SCARPA, T., 2007. *Funding Biomedical Research In the U.S.A.: The Strategic National Importance of NIH Peer Review.* Sigtuna, Sweden: Economic returns of medical research meeting.

SCHMIDT, T., 2005. *Absorptive Capacity – One Size Fits All?A Firm-level Analysis of Absorptive Capacity for Different Kinds of Knowledge.* Discussion Paper No. 05-72. Mannheim, Germany: Centre for European Economic Research.

SCIENCE, INNOVATION AND ELECTRONIC INFORMATION DIVISION, 2008. *Science Statistics: Estimates of Total Spending on Research and Development in the Health Field in Canada, 1996 to 2007.* July 2008 edition. Ottawa, Ontario: Statistics Canada.

SCIENCE-METRIX, 2008-last update, canadian biotechnology patents [Homepage of Science-Metrix], [Online]. Available: http://www.science-metrix.com/eng/projects_2003_Canadian_Biotechnology_Patents.htm [August 20th, 2008].

STATISTICS CANADA, 2008a. *Cancer Incidence in Canada: 2005 and 2006.* 82-231-X. Ottawa, Ontario: Statistics Canada.

STATISTICS CANADA, 2008b. *Science Statistics May 2008 edition.* Catalogue no. 88-001-X. Ottawa, Ontario: Statistics Canada.

STATISTICS CANADA, May 11th, 2007-last update, table 102-0311: potential years of life lost, by selected causes of death and sex, population aged 0 to 74, three-year average, canada, provinces, territories, health regions and peer groups, occasional (51102 series) [Homepage of Statistics Canada], [Online]. Available: http://cansim2.statcan.ca/cgi-win/CNSMCGI.EXE?Lang=E&ArrayId=01020311&Array_Pick=1&Detail=1&ResultTemplate=CII/CII___&RootDir=CII/&TblDetail=1&C2SUB=HEALTH [October 21st, 2008].

STATISTICS CANADA and CANADIAN INSTITUTE FOR HEALTH INFORMATION, 2008. *Health Indicators 2008.* Ottawa, Ontario: CIHI.

STEINBROOK, R., 2006. Private Health Care in Canada. *New England Journal of Medicine,* **354**(16), pp. 1661-1664.

THE ATKINSON CHARITABLE FOUNDATION, 2005. Introducing…The Canadian Index of Wellbeing. *Reality Check - The Canadian Review of Wellbeing,* **5**(1), pp. 1-3.

THE CANADIAN PRESS, October 15th, 2007-last update, wait times for surgery in canada at all-time high: study [Homepage of The Canadian Press], [Online]. Available: http://www.cbc.ca/health/story/2007/10/15/waittimes-fraser.html [August 24th, 2008].

THE COMMON CV SYSTEM, August 28th, 2006-last update, the common CV [Homepage of The Common CV System], [Online]. Available: http://www.commoncv.net/index_e.html [August 25th, 2008].

TIJSSEN, R.J. and VAN LEEUWEN, T.N., 2006. Measuring impacts of academic science on industrial research: A citation-based approach. *Scientometrics,* **66**(1), pp. 55-69.

TREASURY BOARD OF CANADA SECRETARIAT, 2007. *Canada's Performance Report 2006-07: The Government of Canada's Contribution.* Ottawa, Ontario: Treasury Board of Canada Secretariat.

VAN LEEUWEN, T.N. and GRANT, J., 2007. *Bibliometric analysis of highly cited publications of health research in England, 1995-2004.* WR-368-DH. Cambridge, UK: RAND Europe.

VAN RAAN, A.F.J., 2006. Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. *Scientometrics,* **67**(3), pp. 491-502.

VETENSKAPSRÅDET, 2004. *The Medical Research Situation in Sweden.* Stockholm, Sweden: Vetenskapsrådet.

WELLS, R. and WHITWORTH, J., 2007. Assessing outcomes of health and medical research: do we measure what counts or count what we can measure? *Australia and New Zealand Health Policy,* **4**(1), pp. 14.

WILCOX, A.J., 2001. On the importance—and the unimportance— of birthweight. *International Journal of Epidemiology,* **30**(6), pp. 1233-1241.

WOLFSON, M., n.d. The Statistics Canada Graduate Progression Survey.

WOODING, S., 2008. Utilising a payback survey to assess Arthritis Research Council (arc) funding - ongoing research. Cambridge, UK: RAND Europe.

WOODING, S., HANNEY, S., BUXTON, M. and GRANT, J., 2005. Payback arising from research funding: evaluation of the Arthritis Research Campaign. *Rheumatology (Oxford),* **44**, pp. 1145-56.

WOODING, S., HANNEY, S., BUXTON, M. and GRANT, J., 2004. *The Returns from Arthritis Research Volume 1: Approach, Analysis and Recommendations.* MG-251-ARC. Cambridge, UK: RAND Europe.

WORLD HEALTH ORGANIZATION, 2008a-last update, burden of disease statistics: death and DALY estimates for 2002 by cause for WHO member states [Homepage of WHO], [Online]. Available: http://www.who.int/healthinfo/bod/en/index.html [August 21st, 2008].

WORLD HEALTH ORGANIZATION, 2008b-last update, health statistics and health information systems: disability adjusted life years (DALY) [Homepage of WHO], [Online]. Available: http://www.who.int/healthinfo/boddaly/en/ [August 21st, 2008].

Appendix F: Glossary

Accountability: Responsibility and answerability for the use of resources, decisions and/or the results of the discharge of authority and official duties, including duties delegated to a subordinate unit or individual. In regard to programme managers, the responsibility to provide evidence to stakeholders that a programme is effective and in conformity with planned results, legal and fiscal requirements. In organizations that promote learning, accountability may also be measured by the extent to which managers use monitoring and evaluation findings.
Use: UNFPA

Applied Research: Systematic study to gain knowledge or [the] understanding necessary to determine how a recognized and specific need may be met. (As opposed to basic research, which leads to improved general knowledge rather than a practical application).
Use: Alpha Plus[116]

Attribution: Causal link of one event with another. The extent to which observed effects can be ascribed to a specific intervention.
Use: UNFPA

Balanced Scorecard: An analysis technique, developed by Robert Kaplan and David Norton, designed to translate an organization's mission statement and overall business strategy into specific, quantifiable goals and to monitor the organization's performance in terms of achieving these goals.
Use: Oranz Analytics Solutions

Basic Research: Systematic study and investigation undertaken to discover new knowledge, facts or principles. The pursuit of knowledge for the sake of knowledge.
Use: Alpha Plus

Benchmark: Reference point or standard against which performance or achievements can be assessed. Note: a benchmark refers to the performance that has been achieved in the recent past by other comparable organizations, or what can be reasonably inferred to have been achieved in the circumstances.
Use: OECD

Bibliometrics: A field that uses mathematical and statistical techniques, from counting to calculus, to study publishing and communication patterns in the distribution of information.
Use: Diodato 1994

Biomedical Research: Research with the goal of understanding normal and abnormal human functioning at the molecular, cellular, organ system and whole body levels, including development of tools and techniques to be applied for this purpose; developing new therapies or devices that improve health or

---

[116] 'Alpha Plus' is a Canadian Centre of Excellence for resources, standards and innovative use of technology in adult basic education – supports researchers for Deaf, Aboriginal, Francophone and Anglophone communities

the quality of life of individuals, up to the point where they are tested on human subjects. Studies on human subjects do not have a diagnostic or therapeutic orientation.
Use: CIHR2007

Capacity: The knowledge, organization and resources needed to perform a function.
Use: UNFPA

Capacity Development: A process that encompasses the building of technical abilities, behaviours, relationships and values that enable individual groups, organizations and societies to enhance their performance and to achieve their development objectives over time. It progresses through several different stages of development so that the types of interventions required to develop capacity at different stages vary. It includes strengthening the processes, systems and rules that shape collective and individual behaviours and performance in all development endeavours as well as people's ability and willingness to play new developmental roles and to adapt to new demands and situations. Capacity development is also referred to as a capacity building or strengthening.
Use: UNFPA

Case study: A data collection method that involves in-depth studies of specific cases or projects within a program. The method itself is made up of one or more data collection methods (such as interviews and file review).
Use: Statistics Canada

Causal Inference: The logical process used to draw conclusions from evidence concerning what has been produced or "caused" by a program. To say that a program produced or caused a certain result means that, if the program had not been there (or if it had been there in a different form or degree), then the observed result (or level of result) would not have occurred.
Use: Statistics Canada

Clinical Research: Research with the goal of improving the diagnosis and treatment (including rehabilitation and palliation) of disease and injury; improving the health and quality of life of individuals as they pass through
normal life stages. Research on, or for the treatment of, patients.
Use: CIHR 2007

Commercialization: The series of activities undertaken by firms [and institutions] to transform knowledge and technologies (whether developed in Canada or abroad) into new products, processes or services, in response to market opportunities.
Use: Government of Canada

Confounding Factors: The inability to tell between the separate impacts of two or more factors on a single outcome. For example, one may find it difficult to tell between the separate impacts of genetics and environmental factors on depression.
Use: Missouri Institute of Mental Health

Cost-benefit analysis: An analysis that combines the benefits of a program with the costs of the program. The benefits and costs are transformed into monetary terms.
Use: Statistics Canada

Cost-effectiveness analysis: An analysis that combines program costs and effects (impacts). However, the impacts do not have to be transformed into monetary benefits or costs.
Use: Statistics Canada

Counterfactual: The situation or condition which hypothetically may prevail for individuals, organizations, or groups were there no development intervention. Note: Since the counterfactual is a hypothetical state of affairs it cannot be observed but has to be estimated through control group observation, theoretical simulation, and the like. The method of estimating the counterfactual is usually a critical variable in assessments of the validity and reliability of impact evaluations.
Use: OECD, Sida

Covariate: A variable that may affect the relationship between two variables of interest, but is not of intrinsic interest itself. The researcher may choose to control for or statistically reduce the effect of a covariate.
Use: Missouri Institute of Mental Health

Direct Cost: Value of goods and services, expressed in monetary terms, which are directly used in a health intervention. The concept of direct cost is usually used to denote the resources consumed by a health program and may include doctors' time, use of drugs, operation etc. Sometimes it also includes patient's out-of-pocket expenses and resources from other agencies and voluntary bodies.
Use: BDSP

Dissemination: The set of activities by which knowledge about an evaluation is made available to the world at large.
Use: UK Evaluation Society

Effectiveness: A measure of the extent to which a programme achieves its planned results (outputs, outcomes and goals).
Use: UNFPA

Effects: Changes (intended or unintended, positive or negative) resulting directly or indirectly from a public [or private] measure. Effects take into account the outcomes and impacts but not the outputs.
Use: Swiss Federal Office of Public Health

Efficiency: A measure of how economically or optimally inputs (financial, human, technical and material resources) are used to produce outputs.
Use: UNFPA

Evaluation: The systematic and objective assessment of an on-going or completed project, programme or policy, its design, implementation and results. The aim is to determine the relevance and fulfillment of objectives, development efficiency, effectiveness, impact and sustainability. An evaluation should provide information that is credible and useful, enabling the incorporation of lessons learned into the decision-making process of both recipients and donors. Evaluation also refers to the process of determining the worth or significance of an activity, policy, or program. As assessment, as systematic and objective as possible, of a planned, on-going, or completed development intervention. Note: Evaluation in some instance involves the definition of appropriate standards, the examination of performance against those standards, an assessment of actual and expected results and the identification of relevant lessons.

Use: OECD

Evidence based policy: Policy based on results information from scientific research of good quality or other well grounded empirical experience.

Ex-ante Evaluation: An evaluation that is performed before implementation of a development intervention.
Use: UNFPA

Ex-post Evaluation: A type of summative evaluation of an intervention usually conducted after it has been completed. Its purpose is to understand the factors of success or failure, to assess the outcome, impact and sustainability of results, and to draw conclusions that may inform similar interventions in the future.
Use: UNFPA

Feedback: The transmission of findings generated through the evaluation process to parties for whom it is relevant and useful so as to facilitate learning. This may involve the collection and dissemination of findings, conclusions, recommendations and lessons from experience.
Use: OECD

Health: A state of complete physical, mental and social well-being and not merely the absence of disease or infirmity. Health is a resource for everyday life, not the object of living. It is a positive concept, emphasizing social and personal resources as well as physical capabilities. NB: This W.H.O. definition expresses an ideal, which should be the goal of all health development activities. In medicine and in research, health is often understood as an absence of a diagnosed disease or disorder. In the context of health promotion, health is understood as a resource, which permits people to lead an individually, socially and economically productive life in face of ever-changing circumstance.
Use: BDSP

Health literacy: Health literacy represents the cognitive and social skills which determine the motivation and ability of individuals to gain access to, understand and use information in ways which promote and maintain good health.
Use: WHO

Health research: Investigative work undertaken on a systematic and rigorous basis using quantitative and qualitative methods to generate new knowledge that seeks to impact on human physical, social and psychological well being.
Use: Queensland Government

Health Services and Policy Research: Research with the goal of improving the efficiency and effectiveness of health professionals and the health care system, through changes to practise and policy. Health services research is a multidisciplinary field of scientific investigation that studies how social factors, financing systems, organizational structures and processes, health technologies, and personal behaviours affect access to health care, the quality and cost of health care, and ultimately, Canadians' health and well-being.
Use: CIHR 2007

Health technology assessment: An evaluation of the clinical effectiveness, cost-effectiveness, and broader impact of drugs, medical technologies, and health systems, both on patient health and the health care system.
Use: CADTH

Highly Qualified Personnel (HQP): A term that refers to people who have achieved some specified level of educational qualification and thus have a high level of human capital.
Use: Klingbell 2008

Human capital: the knowledge, skills, competencies and attributes embodied in individuals that facilitate the creation of personal, social and economic well-being. HC is an outcome of research.
Use: Klingbell 2008

Impacts: In the context of evaluating health research, the overall results of all the effects of a body of research have on society. Impact includes outputs and outcomes, and may also include additional contributions to the health sector or to society. Impact includes effects that may not have been part of the research objectives, such as contributions to a knowledge based society or to economic growth.
Use: CIHR 2005

Indicator: Quantitative or qualitative factor or variable that provides a simple and reliable means to measure achievement, to reflect the changes connected to an intervention, or to help assess the performance of a development actor. Note: an indicator can also be a measure of an aspect or dimension of change that is unrelated to any particular policy, programme, or project. Governments use social and economic indicators to monitor national developments, and international organizations use indicators in the same way to monitor change regionally and globally.
Use: OECD, Sida

Indirect Costs: Two definitions exist: (1) Indirect costs are all other costs than direct costs, including both indirect costs as defined below in (2) and intangible costs (cost difficult to measure in relation to a disease e.g., fatigue, pain); (2) Indirect costs measure the production loss to society when a human being is unable to produce, for example due to illness or death. The usual way of measuring indirect costs has been to estimate the loss of wage income (approach used in so-called cost-of-illness studies).
Use: BDSP

Inputs: The financial, human, material, technological and information resource provided by stakeholders (i.e. donors, programme implementers and beneficiaries) that are used to implement a development intervention.
Use: UNFPA

Knowledge translation: Funding the actions needed to determine, and mechanisms to support, the actual and potential use of research findings.
Use: MHRC

Logical Framework (Logframe): Management tool used to improve the design of interventions, most often at the project level. It involves identifying strategic elements (inputs, outputs, outcomes, impact) and their causal relationships, indicators, and the assumption or risks that may influence success and failure. It thus facilitates planning, execution and evaluation of a development intervention.
Use: OECD

Metrics: A system of related measures used to assess performance of a program or process and quantify particular characteristics of that program or process.

Outcome: The likely or achieved short-term and medium-term effects of an intervention's outputs. Note: the term can also be defined as the effects that can be directly attributed to an intervention (as opposed to indirect effects) or its effects on the target group (in contrast to its effects on people outside that group). According to a quite different but also common definition an outcome is merely the post-intervention state of the target group or the social conditions that an intervention is expected to have changed. With this definition a change is an 'outcome' even if it is not an effect of the program; even the absence of change is an outcome.
Use: OECD, Sida

Outputs: The immediate tangible results of an activity (e.g. number of papers produced, number of research students).

Partners: The individuals and/or organizations that collaborate to achieve mutually agreed upon objectives. Note: the concept of partnership connotes shared goals, common responsibility for outcomes, distinct accountabilities and reciprocal obligations. Partners may include governments, civil society, non-governmental organizations, universities, professional and business associations, multilateral organizations, private companies, etc.
Use: OECD

Payback: As applied to the logic model is a tool to trace the progress of knowledge and its subsequent utilization, thereby helping to facilitate analysis and consistency in research techniques for data gathering. Also see BUXTON, M.J. and HANNEY, S.R., 1996. How can payback from health services research be assessed? *Journal of Health Services Research Policy*, 1(1), pp. 35-43.

Performance Indicator: A variable that allows the verification of changes in the development intervention or shows results relative to what was planned.
Use: OECD

Performance Measurement: A system for assessing the performance of development interventions against stated goals.
Use: OECD

Performance Monitoring: A continuous process of collecting and analyzing data to compare how well a project, program, or policy is being implemented against expected results. Note: Performance monitoring tends to be descriptive. In order to understand why an intervention has developed as described an in-depth evaluation is often required.
Use: OECD, Sida

Policy: A set of activities, which may differ in type and have different direct beneficiaries, directed towards common general objectives. Policies are not delimited in terms of time schedule or budget.
Use: UK Evaluation Society

Population and Public Health Research: Research with the goal of improving the health of the Canadian population, or of defined sub-populations, through a better understanding of the ways in which social, cultural, environmental, occupational, and economic factors determine health status.
Use: CIHR 2007

Proxy Measure or Indicator: A variable used to stand in for one that is difficult to measure directly.
Use: UNFPA

R&D Personnel: All persons employed directly in research and experimental development (R & D), as well as those providing direct services, such as R&D managers, administrators and clerical staff. Persons providing an indirect service, such as canteen and security staff, should be excluded. R&D personnel comprises researchers, technicians & equivalent staff, and other supporting staff.
Use: UNESCO Institute for Statistics

Researchers: Professionals engaged in the conception or creation of new knowledge, products, processes, methods and systems and also in the management of the projects concerned.
Use: UNESCO Institute for Statistics

Return on Investment (ROI): A measure that evaluates the performance of the Canadian health research enterprise in the categories of Advancing Knowledge; Informing Decision Making; Health Impacts and Economic Benefits.

Scientometrics: Scientometrics [and bibliometrics] are used to measure scientific activities, mainly by producing statistics on scientific publications indexed in databases.
Use: Science-Metrix 2008

Scope: The field of investigation of an evaluation. Typically, this has to be defined from an institutional, temporal and geographical point of view. In addition, one has to identify the key evaluation issues (relevance, efficiency, effectiveness, utility, sustainability) which will be examined.
Use: UK Evaluation Society.

Spinoffs: Firms that have been established by entrepreneurs and have a strong connection to another organization. These entrepreneurs identify discoveries with economic potential that the originating organization chooses not to pursue. The technology may not be pursued because commercialization is outside the mandate of the organization (in the case of universities, hospitals and government labs) or because it is outside the core competence of the organization (in the case of other firms).
Use: Lonmo 2008.

Stakeholders: Agencies, organizations, groups or individuals who have a direct or indirect interest in the development intervention or its evaluation.
Use: OECD

Synthesis: A synthesis is an evaluation or analysis of research evidence and expert opinion on a specific topic to aid in decision-making or help decision makers in the development of policies. It can help place the results of a single study in context by providing the overall body of research evidence. There are many forms of synthesis, ranging from very formal systematic reviews, like those carried out by the Cochrane Collaboration, to informal literature reviews.
Source: CHSRF

Technicians and equivalent staff: Persons with technical knowledge and experience who participate in R&D by performing scientific and technical tasks involving the application of concepts and operational methods, normally under the supervision of researchers.
Use: UNESCO Institute for Statistics

Technology commercialization: Funding to enable researchers who have discovered a potentially useful treatment or tool to work with commercial companies that can bring the discovery into practical (and potentially commercial) use.
Use: MHRC

Technology Transfer: The ability to take a concept from outside the organization (typically from a government or university research programs) and create a product from it.
Use: MIT Sloan School of Management

Technometrics: Technometrics is used to measure specific outputs of R&D, mainly by producing statistics on patents indexed in databases
Use: Science-Metrix 2008

Triangulation: The use of three or more theories, sources or types of information, or types of analysis to verify and substantiate an assessment. Note: by combining multiple data sources, methods, analyses, or theories, evaluators seek to overcome the bias that comes from single informants, single methods, single observers or single theory studies.
Use: OECD

Variance: A descriptive statistic which provides a measure of dispersion. It is obtained by squaring the standard deviation.
Use: UK Evaluation Society

Sources

Alpha Plus Centre, 2008. *Glossary of Research Terms.* [Online]. Available at:
http://alphaplus.ca/pdfs/glossaryresrch.pdf

Banque de Données Santé Publique (BDSP), (n.d.). *Glossaire multilingue.* [Online]. Available at:http://www.bdsp.ehesp.fr/Glossaire/Default.asp

Canadian Agency for Drugs and Technologies in Health (CADTH), 2008. *FAQ.* [Online]. Available at:
http://cadth.ca/index.php/en/hta/faq

Canadian Health Services Research Foundation (CHSRF), (n.d.). *Glossary of Knowledge Exchange Terms as Used by the Foundation.* [Online]. Available at: http://www.chsrf.ca/keys/glossary_e.php

Canadian Institutes of Health Research (CIHR), 2005. *Developing a CIHR Framework to Measure the Impact of Health Research.* [Online]. Available at: http://www.cihr-irsc.gc.ca/e/30418.html

Canadian Institutes of Health Research, 2007. *2007-2008 CIHR Grants and Awards Guide.* [Online]. Available at: http://www.cihr-irsc.gc.ca/e/22630.html#1-A1.1

Diodato V., 1994. *Dictionary of Bibliometrics*. The Haworth Press Inc. Binghampton, New York p. Vii.

Industry Canada, 2006. *People and Excellence the Heart of Successful Commercialization; Volume II: Supporting Material.* [Online]. Available at:
http://www.ic.gc.ca/epic/site/epc-gdc.nsf/vwapj/Volume_II_e.pdf/$FILE/Volume_II_e.pdf

Klingbell, M., 2008. *Measuring Human Capital in the Knowledge Economy.* Centre for Policy Research in Science and Technology, Simon Fraser University.

Lonmo, C., 2008. *Biotechnology spinoffs: transferring knowledge from universities and government labs to the marketplace.* Innovation Analysis Bulletin, Statistics Canada 10 (1).

Missouri Institute of Mental Health, 2002. *Dressed-down Research Terms: A Glossary for Non-researchers.* Program in Consumer Studies and Training, COSP Coordinating Center.

MIT Sloan School of Management, (n.d.). *Keywords & Definitions for IO.* [Online]. Inventing the Organizations of the 21st Century. Available at: http://ccs.mit.edu/21c/iokey.html

On Management Health Group, 2006. *Health Research Funding in Manitoba: A Description and Comparison Across Canada – A Background Paper.* Prepared for the Manitoba Health Research Council. [Online]. Available at: http://www.mhrc.mb.ca/pdfs/MHRC_report_May06.pdf

Oranz AnalyticsSolutions, (n.d.). *Glossary of Terms.* [Online]. Available at:
 http://www.oranz.co.uk/glossary_text.htm

Organization for Economic Cooperation and Development (OECD), 2002. *Glossary of Key Terms in Evaluation and Results Based Management.* The DAC Network on Development Evaluation. [Online]. Available at: http://www.oecd.org/dataoecd/29/21/2754804.pdf

Queensland Government, 2006. *How Does Queensland Health Define Research?* Clinical Practise Improvement Centre. [Online]. Available at:
http://www.health.qld.gov.au/cpic/content/definition.asp

Science-Metrix, 2008. *Scientometrics and Bibliometrics.* [Online]. Available at: http://www.science-metrix.com/eng/methods_scientometrics_t.htm

Science-Metrix, 2008. *Technometrics.* [Online]. Available at: http://www.science-metrix.com/eng/methods_technometrics_t.htm

Statistics Canada, 2002. *Program Evaluation Methods: Appendix 2 – Glossary of Terms.* [Online]. Available at: http://www.tbs-sct.gc.ca/eval/pubs/meth/pem-mep08_e.asp

Swedish International Development Cooperation Agency (SIDA), 2007. *Glossary of Key Terms in Evaluation and Results Based Management.* Department for Evaluation and Internal Audit.

Swiss Federal Office of Public Health, 2005. *Glossary of Evaluation Terms.* Research Policy, Evaluation and Reporting Section.

UNESCO Institute for Statistics, 2007. *Glossary.* [Online]. Available at:
http://www.uis.unesco.org/ev.php?URL_ID=5219&URL_DO=DO_TOPIC&URL_SECTION=201

UK Evaluation Society, 2003. *Glossary of Evaluation Terms.* [Online]. Available at:
http://www.evaluation.org.uk/Pub_library/Glossary.htm

United Nations Population Fund (UNPFA), 2004. *Glossary of Monitoring and Evaluation Terms.* [Online].
Office of Oversight and Evaluation. Available at:
http://www.unfpa.org/monitoring/toolkit.htm

World Health Organization (WHO), 1998. *Health Promotion Glossary.* Division of Health Promotion,
Education and Communications (HPR); Health Education and Health Promotion Unit (HEP).

Appendix G: Methods

This methodology section is set out according to the phases within the timeline of the entire Return on Investment in Health Research assessment. This is done to show the way in which the project has come together, as this is more instructive than simply listing methodologies used in the process.

Phase I: Study Definition

The CAHS Standing Committee on Assessments worked together with the project sponsors to define the precise purpose of the Assessment, its scope, and what it should aim to deliver. This was based on the understanding that there would be a steering panel for the research process made up of experts in health research, research evaluation and impact assessment.

Preparatory Research

Preparatory research included:

- an extensive literature review of methodologies used in evaluating health research (within Canada and worldwide);

- reviewing papers discussing ROI from research; and

- communication with project sponsors.

The literature review found that there were some gaps in the field, particularly in the assessment of social benefits from health research. The review of the literature relied on published and unpublished articles focusing on areas of research methods, returns from research, and overviews of health research landscapes in select countries.

Contact was initiated with sponsors prior to the official launch of the Assessment. Sponsors of the report were contacted by the office of the CAHS President and were asked to identify gaps in the evaluation frameworks of their organizations and what "deliverables" they would like the CAHS report to produce.

After these steps were taken, a project Prospectus was drafted, posted on the CAHS website and distributed to sponsors.

The expert steering panel that would oversee the assessment process was charged with the following tasks:

- to carry out an environmental scan of ROI frameworks and best practises at institutions and agencies both within and outside Canada;
- to consult with policy-makers, the private sector, researchers, funders, other stakeholders and the public about what each value in health research outcomes;
- to propose a framework for measuring the ROI in health research across six domains:
  - o improved health and well being;
  - o benefits to the health care system;
  - o improved decision making and administration;
  - o creation of new knowledge;
  - o increased research capacity for future innovation;
  - o commercial and economic dividends;

- to identify information and human resources that would be required to evaluate ROI on a regular basis in the future; and
- to recognize differences between quantitative and qualitative research and the potential need for different metrics in different research themes.

Phase II: Panel Formation

The steering panel for the Major Assessment was selected by the Standing Committee on Assessments, with input from the Chair of the Assessment. Sponsors and the CAHS Fellowship suggested potential members of the Panel.

In compliance with what is stated in the project Prospectus, the Chair and approximately 25% of the panel members were CAHS Fellows. The remaining panel members comprised both Canadian and international experts.

All panel members were required to sign a Conflict of Interest form and submit their Curriculum Vitaes to the CAHS Board.

Phase III: Panel Deliberation:

Launch of the Assessment

The Assessment was launched at a full day forum held on September 18, 2008. This was part of the agenda of the CAHS 3rd Annual General Meeting (September 17-18, 2007) held in Montreal, Canada.

The Forum introduced the topic of the Assessment to the CAHS Fellowship. Presentations were made by a multi-dimensional group of international and national experts, who represented different stakeholder groups in the Canadian context.

Fellows were broken into break-out groups and asked to brainstorm ideas regarding ROI in one of the following domains: knowledge production; research targeting and capacity; informing policy; health and health sector benefits; and economic benefits. The Academy reconvened, feedback was received from each group and the Forum ended after a discussion and summary of the day. For more details, see the CAHS Forum Summary (http://www.cahs-acss.ca/e/pdfs/ROI_ForumSummary2007.pdf).

The CAHS Forum Summary was provided to the steering panel to provide them background and insight into the scope of the project.

Phase IV: Assessment process

Methodologies

Similar to previous studies in the field of health research evaluation, this project used a range of methods to triangulate research and strengthen findings. The methods used for this study included continuous scanning of ROI literature, key informant interviews, and the commissioning of position papers.

The initial portion of the project strove to create an understanding of the Canadian health research context – to benefit the international panellists on the steering panel, as well as to inform the background of the report. The purpose of this environmental scan was to identify current evaluation methods in place in Canada and identify gaps in these evaluation methods.

The methodologies employed by the steering panel:

- built on existing performance measurement work;
- consisted of a variety of approaches and measures in order to satisfy the wide range of stakeholders who had an interest in this project;
- took into consideration the measurement of both long and short term impacts of research;
- separated economic (commercial) return from social return; and
- involved interactions with sponsor funding agencies and other organizations interested in the same subject material in order to avoid duplication and to tailor the targets of the report appropriately.

This report is based on research carried out between September 2007 and September 2008, using qualitative research methods. Articles, working papers and websites were regularly consulted and were generally provided by organizations or individuals the ROI staff spoke with.

As stated above, a preliminary literature scan was completed by the office of the CAHS president prior to the official launch of the Assessment. No case studies were carried out in this project.

Panel Communication

The ROI staff worked together with the panel to keep them regularly updated of the progress of the report.

The content of this report was discussed at three face-to-face meetings of the steering panel as well as through teleconferences and email.

Additionally, at the half way point of the project, the steering panel was broken into four working groups. These groups focused on specific sub-topics within the report.

Literature review

Throughout the process of assessment, literature around the subject of research evaluation was reviewed. Using initial key references identified through the steering panel, CAHS and other research experts, a snowballing technique of literature identification through references was combined with searches for relevant literature through a number of search engines including Google, Google Scholar, PubMed, and Web of Science. As well as these search engines, specific organization websites were searched using their embedded search engines; these included Canadian government departments, federal and provincial research funders, and international organizations such as the OECD and WHO. References were housed in the web-based reference software 'RefWorks' (ProQuest 2001).

Key Informant Interviews – Semi-structured Interviewing

Semi-structured interviews were a large component of the Assessment. Interviews were carried out by the ROI Staff (and occasionally ROI Panel Members) and respondents.

The Panel and staff engaged in two types of interviews:

- Sponsor interviews

- External expert interviews.

Sponsor Interviews

Additional contact was made with the sponsors of the Assessment. Individuals interviewed were high-level employees (sometimes the CEO or Director) of an organization, or an individual in charge of the Research Evaluation branch within their organization/institution.

All sponsors were contacted via email through the ROI Office and were asked for an interview. A summary of the Assessment, the project Prospectus, and additional supporting documents were included in the request emails. Out of 23 sponsors, the ROI staff was successful in obtaining feedback from 19.

The majority of the communication with sponsors was through teleconferences. All calls were recorded and notes were taken during interviews. All interviewees were asked permission to record the call. Recordings were used only for staff/panel purposes and were not distributed outside of the ROI office.

Questions for semi-structured interviews consisted of a list of 8 questions, which was decided and agreed on by the panel prior to the commencement of the calls. Sponsors were sent the list of questions prior to a call. Interview schedules were tailored according to the availability of sponsors, and secondly of panellists and staff.

This set of interviews was extremely important, as it confirmed what the sponsors were looking for in the report. Information gained from these interviews was not used directly in the report, rather, it served to inform panel members and provide background to the project. Additionally, the findings from these interactions with sponsors were shared and discussed with the steering panel at its second face-to-face meeting.

External Expert Interviews

At the first and second face-to-face meetings, the steering panel identified a range of potential individuals with expertise in certain areas that they felt would help inform the project. A preliminary list of over 50 individuals was made; a number which was reduced through discussions with the panel once the full list had been sub-divided into experts within specific pillars of health research. The aim was to interview at least 3 experts within each pillar, although their specific expertise may not be in the full scope of research within that pillar.

16 individuals from a variety of backgrounds were interviewed. These interviews followed a similar structure to sponsor interviews: potential interviewees were contacted via email and sent a project summary, the Prospectus, and other supporting documents.

All calls were recorded and notes were taken during interviews. All interviewees were asked permission to record the call. Recordings were used only for staff/panel purposes and were never distributed outside of the ROI office. Questions were tailored for each interviewee and sent prior to a call. Interview schedules were tailored according to the availability, primarily of external experts, and secondly of panellists and staff.

This set of interviews was extremely important, as it provided direction and insight into particularly thorny issues of the report. Like the sponsor interviews, information gained from these interviews was not used directly in the report, rather, it served to inform the report, the panel members and provide background to the assessment.

Once the set of interviews concluded, ROI staff performed a conceptual cluster analysis of key themes from the interviews, using qualitative analysis of interview notes and Hexie clustering of the key points identified from the interviews. This clustering provided the underlying structure for the analysis of the interview findings. Hexie clustering is an analysis technique often used in public policy research and

involves the noting of findings on hexagonal post-it notes that can then be rearranged to form clusters along different thematic lines (Wooding, Scoggins et al. 2005). These themes were summarised in a briefing document sent to the steering panel.

The outcome of these interviews complemented research undertaken by ROI staff, provided a full contextual background and informed the selection of metrics and a framework. They aided the identification of what should be considered in the report in the view of different stakeholders in the system.

Commissioned Papers

The steering panel and Chair identified several "gaps" in knowledge of the subject. In response to this, the Chair commissioned six commissioned papers from external experts. The use of commissioned papers is common practice in other countries where bodies equivalent to the CAHS produce reports, for example the Institute of Medicine in the USA.

Three papers were commissioned on three of the four pillars of CIHR research. The paper on Pillar II (Clinical Research) was written by Dr. Ralph Meyer. The paper on Pillar III (Health Services and Policy Research) was written by Steven Lewis, Patricia Martens, and Louis Barre. The paper on Pillar IV (Population and Public Health Research) was written by Dr. Alan Shiell and Erica Di Ruggiero. Pillar I, basic biomedical research, is probably the best understood of the pillars in terms of 'payback', it is also the most researched for impact analyses. For this reason, there was no paper commissioned in Pillar I.

Three additional papers were commissioned on topics that steering panel members identified as important additional areas of focus for this assessment. The paper on the role Ethics in Health Research Evaluation was written by Drs. Michael McDonald and Bartha Knoppers. The paper on Assessment of impact at the Meso Level was written by Dr. Jerald Hage. The paper on the Public Perspective of Health Research was written by Andre Picard.

Phase V: External Review

After the third and final meeting of the steering panel, ROI staff modified the draft report as necessary. Shortly thereafter, the draft report was sent to the Standing Committee on Assessments, who forwarded it to an External Review Committee (chosen in advance by the Standing Committee). The identity of the External Review Committee was kept confidential.

The steering panel and ROI staff then evaluated the report based on the recommendations from External Review and made necessary changes.

The approval and acceptance of the final version of the report was made by the CAHS Council.

References

PROQUEST, 2001. RefWorks. January 2008 edn. Bethesda, MD: CSA.

WOODING, S., SCOGGINS, A., LUNDIN, P. and LING, T., 2005. *Talking Policy: An examination of public dialogue in science and technology policy.* Cambridge, UK: RAND Europe.

Appendix H: External Interviewees

As part of the process of understanding the main issues surrounding ROI in health research, we interviewed a number of stakeholders in the process. These included experts in particular aspects of evaluation, researchers in particular pillars, and researchers from a variety of research organizations. The list of external interviewees is below, with their organizations identified.

Pillar I Informants

Douglas Barber: Distinguished Professor in Residence, McMaster University

•   Former President & CEO, Gennum Corporation; Past Vice Chair, Ontario Science and Innovation Council; Past Member, Ontario Postsecondary Education Quality Assessment Board; Director, Canadian Academy of Engineering.

Ilse Treurnicht: CEO, MaRS Discovery District

•   Former president and CEO of Primaxis Technology Venture; former entrepreneur with senior management roles in a number of start-up companies; Director of MaRS, Primaxis, BTI Photonic Systems, Optimer Photonics, the Toronto Venture Group, BioDiscovery Toronto and the Canadian Institute for Advanced Research (CIAR).

Pillar II Informants

John Cairns: Professor, Department of Cardiology, University of British Colombia

•   Dean Emeritus, University of British Colombia; Former Chair of Medicine, McMaster University; Project Leader of the CIHR Clinical Research Initiative (strengthening Canada's endeavour in clinical research).

Stefan Ellenbroek: Research Policy Advisor at Leiden University Medical Centre (Netherlands)

•   Manages the introduction of the 'societal impacts framework' for assessing research impacts at LUMC.

Pillar III Informants

Greg Webster: Director of Research and Indicator Development Canadian Institute for Health Information (CIHI)

•   Author of 'Indicators for Primary Health Care' from CIHI amongst other publications

Pillar IV Informants

Wendy Baldwin: Director, Poverty, Gender, and Youth Program (Population Council)

•   Previous deputy director for extramural research (NIH); Served on committees with the NAS, AAAS, Department of Health and Human Services (USA); Works with WHO.

Egon Jonsson: CEO, Institute of Health Economics

•   Professor of Public Health Sciences at the University of Alberta; Former Professor of Health Economics at the Karolinska Institute, Stockholm, Sweden; Member of NAS (US) ; Editor of the International Journal of Technology Assessment in Health Care; Worked with WHO Euro to

establish a Health Evidence Network; Health Policy Advisor at the Ministry of Health, Hanoi, Vietnam.

Noralou P.Roos: Canada Research Chair in Population Health Research

• Professor in the Department of Community Health Sciences, University of Manitoba; Founding Director of the Manitoba Centre for Health Policy; Former member of the Prime Minister's Health Forum; Established the Population Health Research Data Repository at the Manitoba Centre for Health Policy and Evaluation.

Sharon Manson Singer: President, Canadian Policy Research Networks

• Has held deputy minister positions in the BC government; Adjunct professor at the School of Public Administration (University of Victoria); has served as an expert advisor to all levels of government in Canada.

Public Perspective/Other

Teren Clarke: Executive Director, Canadian Paraplegic Association (Alberta Division)

• Former National Director of Programs and Services, Muscular Dystrophy Canada.

Cheryl L. Koehn: Founder and President, Arthritis Consumer Experts

• Principal advisor for Canadian Arthritis Network (CAN); Author on arthritis issues for patients.

Greg Tassey: Senior Economist, National Institute of Standards and Technology (NIST)

• Author of multiple publications and books on R&D trends and associated policy implications.

Muhajarine Nazeem: Research leader, Healthy Population Domain - Canadian Index of Wellbeing

• Professor and Chair in Community Health and Epidemiology, University of Saskatchewan; leads Saskatchewan Population Health and Evaluation Research Unit's Healthy Children research area.

Carol Dahl: Director, Global Health Discovery program, Gates Foundation

• Former vice president for Strategic Partnerships at Biospect Inc.; founding director of the Office of Technology and Industrial Relations at the National Cancer Institute (part of NIH); former program director at the National Center for Human Genome Research.

Appendix I: Prospectus for a Major Assessment – The Return on Investments in Health Research: Defining the Best Metrics

Prepared by the Canadian Academy of Health Sciences

December 2007

The Return on Investments in Canadian Health Research –The Situation

Investments in health research have increased significantly across Canada over the past decade. Naturally, and justifiably, with these greater investments come increased expectations. In addition, the widening diversity of stakeholders engaged in and/or supporting health research has led to a broader range of anticipated outcomes. These expectations include: 1] better health; 2] greater life expectancy; 3] translation of research findings into improvements in quality of life; 4] informed public policy on health related issues across the full spectrum of government and private sector activity; 5] new commercial opportunities within and beyond Canadian borders; 6] increased attraction of the next generation to pursue careers in health research and the health sector; 7] a better 'state of readiness" for the unexpected threats to health that inevitably develop in the contemporary world.

In parallel with these expectations, a confluence of factors has placed intense focus on understanding what return our society receives for the investments made in health research. Some of these include:

- lack of public understanding of the value of research and its applicability to current issues in health care at a time of unsurpassed concern about accessible, affordable, high quality health care in a publicly funded system;

- failure to adequately measure the benefits of fundamental and applied health research and to properly convey them in a meaningful fashion to policy-makers and the public;

- an increasingly common view that health care (and by association, health research) is a cost-driver consuming an ever greater share of provincial resources at the expense of other sectors;

- growing general concern about expenditure accountability in the aftermath of widely publicized instances of misuse in both the public and private sectors in Canada and abroad;

- lack of consensus on how and when to best evaluate return on research expenditures;

- specific questions from policy makers about tangible results attributable to recent increases in public investment in health research through the Canadian Institutes of Health Research, the Canada Foundation for Innovation and the Canada Research Chairs program;

- uncertainty about the appropriateness of Canada's expenditures on health research versus those of analogous contributions in other industrialized countries;

- a need to acquire appropriate evidence to assist in striking the right balance between funding of investigator-initiated "discovery" health research and targeted "strategic" health research;

- a decline in the number of health professionals pursuing health research careers at a time when the "greying" of current investigators is likely to lead to a major decline in research personnel;

- mounting pressure on innovation as the primary avenue for improving Canadian productivity and standard of living in the knowledge based economy of the 21st century;

- the need for a robust multi-dimensional measurement framework that addresses the increasingly complex, multi-sectoral impacts of health research spanning:

  - improved health and well being

  - benefits to the health care system

  - improved decision making and administration

  - creation of new knowledge

  - training of the next generation of researchers for future innovation

  - commercial and economic dividends

Potential Scope

The scope and deliverables of the Assessment will be based on joint agreement between CAHS and the Sponsors. The general intention is to propose a clear menu of metrics by which return on investments in health research in Canada can be measured. It is understood that different Sponsors will possess a varied spectrum of interest about different metrics.

The procedures to conduct the Assessment will be determined by the Assessment Panel and may include receipt of written submissions, open and closed meetings of the Panel, and forums involving the Panel, Sponsors and leading authorities within and outside of Canada.

The final report may contain some or all of the following:

- environmental scans of return on investment frameworks and best practices at institutions and agencies both within and outside of Canada; this includes incorporation of previous work conducted by CIHR , and casting international work from the UK , Australia and other countries into the Canadian context

- consultations with policy-makers, the private sector, researchers, funders, other stakeholders and the public about what each value in health research outcomes

- a framework for measuring the return on investments in health research across the six domains listed below:

  - improved health and well being

  - benefits to the health care system

  - improved decision making and administration

  - creation of new knowledge

  - increased research capacity for future innovation

- commercial and economic dividends
- identification of the information resources and human resources that would be required to evaluate returns on investment on a regular basis in the future
- a recognition of the differences in quantitative and qualitative research and potential need for differing metrics in different research themes
- other elements deemed relevant by the stakeholders

Tentative Workplan

Phase I: Study Definition:

The CAHS Standing Committee on Assessments together with the Assessment Sponsors will define the precise nature of the question, the scope of the Assessment and the assessment deliverables.

Phase II: Panel Formation:

All Sponsors, the CAHS Fellowship, other interested parties and the public will be invited to suggest potential members of the Assessment Panel. The Standing Committee on Assessments will propose a membership list of the Assessment Panel to the CAHS Board. The Chair and approximately 25% of the members will be Fellows of CAHS. The remaining 75% of members will be selected from the best Canadian and international experts in the field and will include public representation.

The proposed panel will be posted on the CAHS web-site for comment and suggestions prior to finalization. Final approval of the Assessment Panel will rest with the CAHS Board.

Phase III: Panel Deliberation:

The Panel together with professional/ support staff will conduct their work. This will include environmental scanning, receipt of written submissions by interested parties, open hearings with presentations from interested parties, closed meetings and deliberations. Consideration will be given to launching the assessment process with a Major Forum involving leading international experts to which the Sponsors will be invited.

Phase IV: External Review:

A draft report will be received by CAHS and forwarded to an External Review Committee selected by the Standing Committee on Assessments. Sponsors will again be invited to suggest members of the External Review Committee. The Assessment Panel will subsequently evaluate its report based on recommendations from External Review. Approval and acceptance of the final report will rest with CAHS Council.

Phase V: Dissemination:

The final report will be distributed widely in printed format and posted on the CAHS web site. Other methods of dissemination, based on prior agreement with the Sponsors, will be utilized. These may include presentations, town hall meetings, non-print media, etc. in order to maximize the impact and uptake of the recommendations.

Budget

Estimated range: $500,000 to $600,000

The final budget will depend on scope and variable costs such as number of meetings and hearings. The final budget will be agreed upon in advance through written contract between CAHS and the Sponsors.

It is anticipated that the funding costs would be shared among a large number of institutions and agencies heavily impacted by this complex set of issues, leading to a relatively low cost per individual sponsor.

Assessment Sponsors

Major Sponsors

• Canadian Health Services Research Foundation (CHSRF)
• Canadian Institutes of Health Research (CIHR)
• Canada's Research Based Pharmaceutical Companies (Rx & D)
• Public Health Agency of Canada (PHAC)

Sponsors

• Alberta Heritage Foundation for Medical Research (AHFMR)
• Association of Canadian Academic Healthcare Organizations (ACAHO)
• Association of Faculties of Medicine of Canada (AFMC)
• BIOTECanada
• Canadian Agency for Drugs and Technologies in Health (CADTH)
• Fonds de la recherche en santé du Québec (FRSQ)
• Government of Ontario, Ministry of Research and Innovation; Ministry of Health and Long-Term Care
• Heart and Stroke Foundation of Canada (HSFC)
• Manitoba Health Research Council (MHRC)
• Michael Smith Foundation for Health Research (MSFHR)
• National Cancer Institute of Canada (NCIC)
• Nova Scotia Health Research Foundation (NSHRF)
• Ontario Neurotrauma Foundation (ONF)
• Saskatchewan Health Research Foundation (SHRF)
• Western Economic Diversification Canada (WD)

Contributors

• Canada Foundation for Innovation (CFI)
• Canadian Association of Schools of Nursing (CASN)
• Canadian Medical Association (CMA)
• Canadian Nurses Association (CNA)
• Canadian Nurses Foundation (CNF)
• Newfoundland & Labrador Centre for Applied Health Research (NLCAHR)
• Research Canada

About the Canadian Academy of Health Sciences

The Canadian Academy of Health Sciences (CAHS) is comprised of approximately 200 Fellows who have attained the highest levels of academic and professional accomplishment in their respective fields. CAHS is not an advocacy group but rather an organization comprised of individuals from diverse backgrounds who have agreed to volunteer their time and expertise to participate in assessments of crucial health- and biomedical related issues affecting the lives of all Canadians.

The objectives of CAHS are to:

1. Serve as a credible, expert, independent assessor of science & technology issues relevant to health of Canadians

2. Support the development of timely, informed & strategic advice on urgent health issues

3. Facilitate development of sound & informed public policy on these issues

4. Enhance understanding of health-related science & technology issues affecting the public by transmitting results of assessments & providing opportunities for public discussion

5. Provide a single authoritative & informed voice for the health science communities

6. Monitor global health issues to enhance Canada's state of readiness for the future

7. Represent Canadian health sciences internationally & liaise with international academies to enhance understanding and potential collaborations

Remarkably, until now, Canada has been unique in not having this type of resource as compared with many other countries such as the United States, France, the Netherlands, and the United Kingdom. Both the U.S. Institute of Medicine and the U.K. Academy of Medical Sciences are interdisciplinary organizations that respond to questions and issues put to them from a variety of sources: government, national non-governmental organizations, industry, academia and major research organizations. Below are some of the reports that the Institute of Medicine in the U.S. has produced after careful study and analysis that have had a meaningful impact on all aspects of health:

- To Err is Human: Building a Safer Health System (1999)

- Stem Cells and the Future of Regenerative Medicine (2001)

- Crossing the Quality Chasm: A New Health System (2001)

- Who will keep the Public Healthy? Educating Public Health Professionals for the 21st Century (2002)

- Preventing Childhood Obesity: Health in the Balance (2004)

CAHS Fellows

Distinguished Fellows
Monique Bégin
John Evans

Emeritus Fellows
Tasso P. Anastassiades
Patricia A. Baird
Manuel Buchwald
Donald Calne
Serge Carrière
Ray Chiu
Anthony Chow
Michel Chrétien
Richard Cruess
Henry Dinsdale
Henry Friesen
Jacques Genest
J. Richard Hamilton
James C. Hogg
Martin Hollenberg
George Karpati
Wilbert J. Keon
Otto Kuchel
Bernard Langer
Peter Macklem
Ernest A. McCulloch
T. Jock Murray
J. Fraser Mustard
Reginald A. Nadeau
Arnold Naimark
Barry Pless
Domenico Regoli
Claude Roy
Robert B. Salter
Michael J. Sole
Matthew Spence

Current Fellows
Albert J. Aguayo
William Albritton
Aubie Angel
Jack Antel
Stephen Archer
Paul W. Armstrong
Francois A. Auger

Lorne Babiuk
Michael Baker
Penny Ballem
Morris Barer
Ronald Barr
Renaldo Battista
Francoise Baylis
Alain Beaudet
Michel Bergeron
Howard Bergman
Alan Bernstein
Charles N. Bernstein
Allan Best
John Bienenstock
Joan Bottorff
Michel Bouvier
M. Ian Bowmer
Bernie H Bressler
Donald Brooks
John Brosnan
Helen Burt
John Cairns
André M. Cantin
S. George Carruthers
Neil R Cashman
Carol Cass
Vincent Castellucci
Timothy Caulfield
Benoit Chabot
Thomas MS Chang
Sylvain Chemtob
Davy Cheng
Harvey Chochinov
M. Thomas Clandinin
John Conly
Andre-Pierre
Contandriopoulos
Alastair Cribb
Max Cynader
Abdallah Daar
Dale Dauphinee
Jean Davignon
Dave Davis
Jacques de Champlain

Lesley Degner
John Denstedt
Johanne Desrosiers
Naranjan Dhalla
John Dirks
Ian Dohoo
Allan Donner
Diane Doran
James Dosman
Francine Ducharme
Andreé Durieux-Smith
Nancy Edwards
Mostafa Elhilali
Mary Ensom
John Esdaile
Carole Estabrooks
Robert Evans
John Fairbrother
Thomas Feasby
Howard H Feldman
Diane Finegood
B. Brett Finlay
William A. Fisher
Jean-Claude Forest
Pierre-Gerlier Forest
Cyril Frank
John Frank
William Fraser
Abraham Fuks
Grant Gall
Nicole Gallo-Payet
John H.V. Gilbert
Phil Gold
Larry Goldenberg
Harry L. Goldsmith
David Goltzman
Avrum Gotlieb
Céline Goulet
Paul Grand'Maison
Jean Gray
Jeremy Grimshaw
Ronald D. Guttmann
Harvey Guyda
Carlton Gyles

Vladimir Hachinski
Antoine Hakim
Judith Hall
Phillip Halloran
Pavel Hamet
David F. Hardwick
Susan Harris
David A. Hart
David Hawkins
Michael Hayden
Rejean Hebert
Robert Hegele
Carol Herbert
Clyde Hertzman
Philip Hicks
Wayne Hindmarsh
Ellen Hodnett
Murray W. Huff
Alex Jadad
Jack H. Jhamandas
Yves Joanette
Joy Johnson
Celeste Johnston
Jawahar (Jay) Kalra
Norah Keating
Nuala Kenny
Kevin Keough
Terry Klassen
Bartha Knoppers
Michael S. Kramer
Stanley P. Kutcher
Fernand Labrie
Jean-Claude Lacaille
André Lacroix
Andreas Laupacis
Mary Law
Yvonne Lefebvre
Franco Lepore
Peter C.K. Leung
Wendy Levinson
Peter Liu
David Locker
Jonathan Lomas
Donald Low
James Lund
Renee F. Lyons
Noni MacDonald
Stuart MacLeod

Andrew Macnab
Annette Majnemer
Paul Man
G. B. John Mancini
Karen Mann
Thomas Marrie
James G. Martin
Renée Martin
S. Wayne Martin
Anne Martin-Matthews
Christopher McCulloch
Grant McFadden
Patrick McGrath
Roderick McInnes
Bruce McManus
John McNeill
Graydon Meneilly
Jose Menezes
Richard Morisset
Barbara Morrongiello
Janice Morse
Jean-Marie Moutquin
David S. Mulder
Bruce Murphy
Louise Nasmith
Stanley Nattel
C. David Naylor
Lindsay Nicolle
Jeff Nisker
Hugh O'Brodovich
Linda O'Brien-Pallas
Annette O'Connor
Chris Overall
Peter Pare
André Parent
Patrick S. Parfrey
Hélène Payette
Jean-Pierre Pelletier
Eliot Phillipson
Grant Pierce
Roger Pierson
Martha Piper
Frank Plummer
Louise Potvin
Barry I. Posner
John F Prescott
Dorothy Pringle
Remi Quirion

Raymond Rajotte
Eugenio A. Rasio
Pamela Ratner
Marie-France Raynault
Jeffrey Reading
Paul Rennie
Richard Reznick
Carol Richards
Richard Riopelle
Kenneth Rockwood
Allan Ronald
Irving Rootman
Lawrence Rosenberg
David Rosenblatt
Walter Rosser
Serge Rossignol
Ori D. Rotstein
Guy Rouleau
Jean-Lucien Rouleau
Rima Rozen
Ellen Rukholm
Guy Sauvageau
Martin T. Schechter
Ernesto Schiffrin
Rafick Sékaly
Barry Sessle
Robert S. Sheldon
Susan Sherwin
Melvin Silverman
Jack Siemiatycki
Jacques Simard
Robert D. Sindelar
Peter Singer
Bhagirath Singh
Emil Skamene
Ingrid Sketris
Harvey Skinner
Arthur Slutsky
Eldon R. Smith
Bonnie Stevens
Terrance P. Snutch
J David Spence
Heather K. Spence-
    Laschinger
George Steiner
Miriam Stewart
Sherry Stewart
Donald Stuss

Roger A. Sutton
Robyn Tamblyn
Jean-Claude Tardif
Charles H. Tator
Sally Thorne
Aubrey Tingle
Johanne Tremblay
Richard Tremblay
Jack Tu
Peter Tugwell
Jacques Turgeon
Jeffrey Turnbull
D. Lorne Tyrrell
Jack Uetrecht
Patrick Vinay
Mark A. Wainberg
Peter Walker
Keith Walley
Mamoru Watanabe
Donald Weaver
Charles Weijer
Jeffrey I. Weitz
Catharine Whiteside
Douglas Wilson
Michael Wolfson
Sharon Wood Dauphinee
Donald Woods
Ronald G. Worton
Salim Yusuf

Honorary Fellow
Kenneth Fung