Canadian Academy of Health Sciences
ASSESSMENT REPORT

MAKING AN IMPACT

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



Report of the Panel on the Return on Investments in Health Research
January 2009

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

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/

CAHS wishes to thank the sponsors of this assessment without whom this report would not have been possible. This study was supported by A) 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) B) 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 & 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) and C) 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) and Research Canada. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the view of the organizations or agencies that provided support for this project.

ISBN: 978-0-9811589-0-7

Printed in Canada

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

Published 2009 by the Canadian Academy of Health Sciences

774 Echo Drive, Ottawa, Ontario, Canada K1S 5N8

Tel/Tél 613 260 4174 | Fax Téléc 613 730 1116 E-mail/ courriel: cahs@rcpsc.edu

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/

The Canadian Academy of Health Sciences

The Canadian Academy of Health Sciences was founded in 2004 as a non-profit, charitable organization composed of elected Fellows from diverse disciplines both within and external to the health sector. It is both an honorific membership organization and a policy research organization. The Fellows are elected to the Academy by a rigorous peer-review process that recognizes demonstrated leadership, creativity, distinctive competencies and a commitment to advance academic health science. In this way, the Canadian Academy of Health Sciences joins the Royal Society of Canada, which has long honoured Canadians for their accomplishments in the natural sciences, social sciences and humanities, and the Canadian Academy of Engineering, which recognizes achievements in the engineering sciences.

A unique aspect of the Canadian Academy of Health Sciences, relative to international comparators, is that its Fellows span the full breadth of academic health science from fundamental science to social science and population health; they also represent the full spectrum of health disciplines, including nursing, dentistry, veterinary medicine, rehabilitation sciences, pharmaceutical sciences, medicine and related fields such as psychology, health law, ethics and health economics. As a result, the Canadian Academy of Health Sciences is truly ecumenical and well disposed to provide a holistic perspective on health-related subjects.

The objectives of the Canadian Academy of Health Sciences are to:

- serve as a credible, expert and independent assessor of science and technology (S&T) issues relevant to the health of Canadians;
- support the development of timely, informed and strategic advice on urgent health issues;
- support the development of sound and informed public policy related to these issues;
- enhance understanding of S&T issues affecting the public good by transmitting the results of assessments and providing opportunities for public discussion of these matters;
- provide a single authoritative and informed voice for the health science communities;
- monitor global health related events to enhance Canada's state of readiness for the future.
- represent Canadian health sciences internationally and liaise with like international academies to enhance understanding and potential collaborations on matters of mutual interest.

In short, the Canadian Academy of Health Sciences provides “scientific advice for a healthy Canada”. The challenges facing governments at all levels, institutional and professional leaders in the health system, the non-governmental and business sectors, and the public in regard to health and the health care system are complex and daunting. Such issues require careful and thoughtful analysis that is not only expert, but also unbiased and independent of vested interests and agendas. They call for an objective weighing of the available scientific evidence at arm’s length from political considerations and with a focus on the public interest.

The process of the Academy’s work is designed to assure appropriate expertise, the integration of the best science and the avoidance of bias and conflict of interest, the latter being a frequent dynamic that

confounds solutions to difficult problems in the health sector. While those organizations that sponsor assessments have input into framing the study question(s), they cannot influence the outcomes of an assessment or the contents of a report. Academy reports undergo extensive review and evaluation by external experts who are anonymous to the panel, and whose names are revealed only once the study is released. Final approval for release and publication of an Academy Report rests only with the Board of CAHS.

Panel Members and Staff

Assessment Panel Members

- Cyril Frank, MD (Chair), McCaig Professor of Joint Injury and Arthritis Research; Executive Director of the Alberta Bone and Joint Health Institute; and Professor, Division of Orthopaedics University of Calgary, Calgary, AB, Canada
- Renaldo Battista, MD, MPH, ScD, Professor and Director of the Department of Health Administration, Université de Montréal, Montreal, QC, Canada
- Linda Butler, Fellow and Head, Research Evaluation and Policy Project, Australian National University, Canberra, Australia
- Martin Buxton, BA, Professor, Health Economics, Brunel University, United Kingdom
- Neena Chappell, PhD, Canada Research Chair; Social Gerontology, Professor of Sociology and Centre on Aging, University of Victoria, Victoria, BC, Canada
- Sally C. Davies, Director General, Research and Development, Department of Health and National Health Service, United Kingdom
- Aled Edwards, PhD, Banbury Professor, Banting and Best Department of Medical Research, University of Toronto, Toronto, ON, Canada
- Chris Henshall, PhD, Pro-Vice-Chancellor, External Relations, University of York, York, England, United Kingdom
- Yann Joly, LLB, LLM, project manager, Centre de recherche en droit public Université de Montréal, Montreal, QC, Canada
- Gretchen Jordan, PhD, Principal Member of Technical Staff, Science and Technology Strategic Management Unit, Sandia National Laboratories, Department of Energy, Washington, DC, United States
- Terence Kealey, MB, BS, PhD, Vice Chancellor & Clinical Bio-Chemist, University of Buckingham, Buckingham, United Kingdom; author *The Economic Laws of Scientific Research*
- Michael C. Wolfson, PhD, Assistant Chief Statistician, Analysis and Development, Statistics Canada, Ottawa, ON, Canada
- Steven H. Woolf, MD, MPH, Professor, Departments of Family Medicine, Epidemiology, and Community Health, Virginia Commonwealth University, Richmond, VA, United States

After approval for membership to the Assessment Panel by the Canadian Academy of Health Sciences' Standing Committee on Assessments, all members were required to declare in writing any potential conflicts of interests. These are available for review on request. The Committee was responsible for oversight and internal review of this Assessment.

ROI Assessment Staff

- Rhonda Kennedee, Meeting and Events Coordinator
- Linda Marchuk, Research Associate
- Edward Nason, Health Research Evaluation Analyst and Writer
- Larissa Sommerfeld, Research Assistant

External Reviewers

- Joseph B. Martin. MD, PhD, Edward R. and Anne G. Lefler Professor of Neurobiology and former Dean of the Harvard Faculty of Medicine, Harvard Medical School, Boston, MA, USA
- The Honourable Michael J. L. Kirby, MA, PhD, LLD(Hon), Chair, Mental Health Commission of Canada, Senator (former), Ottawa, ON, Canada
- John W. Frank, MD, CCFP, MSc, FRCP(C), Director, Scottish Collaboration for Public Health Research and Policy, MRC Human Genetics Unit, Edinburgh, United Kingdom

CAHS' Standing Committee on Assessments

- Andreas Laupacis, MD (Chair), Executive Director, Li Ka Shing Knowledge Institute of St. Michael's Hospital; Professor, Faculty of Medicine, University of Toronto, Toronto, ON, Canada
- John Allen Cairns, MD, Professor of Medicine and Dean Emeritus, University of British Columbia, Vancouver, BC, Canada
- Timothy Caulfield, LLM, Canada Research Chair in Health Law and Policy; Professor, Faculty of Law and School of Public Health; Senior Health Scholar with the Alberta Heritage Foundation for Medical Research and Research Director, Health Law Institute, University of Alberta, Edmonton, AB, Canada
- André-Pierre Contandriopoulos, PhD, Professeur Titulaire, Département d'Administration de la santé, Université de Montréal, Montréal, QC, Canada
- Alastair Cribb, DVM, PhD, Dean, Faculty of Veterinary Medicine, University of Calgary, Calgary, AB, Canada
- Jean Gray, CM, MD, FRCPC, LLD, DSc, Professor Emeritus, Dalhousie University, Halifax, NS, Canada
- Pavel Hamet, MD, PhD, Canada Research Chair, Predictive Genomics; Chief, Gene Medicine Services, Centre de recherche Centre hospitalier de l'Université de Montréal (CHUM), Montréal, QC, Canada
- Dorothy Pringle, RN, BScN, MS, PhD, Professor Emeritus, Faculty of Nursing, University of Toronto, Toronto, ON, Canada
- Matthew Spence, OC, MD, PhD, Retired President and Chief Executive Officer, Alberta Heritage Foundation for Medical Research, Edmonton,, AB, Canada
- Peter S. L. Tugwell, MD, Canada Research Chair in Health Equity; Director, Centre for Global Health (Institute of Population Health); Professor of Medicine & Epidemiology, University of Ottawa, Ottawa, ON, Canada
- Sharon L. Wood Dauphinee, PhD, PT, Professor, McGill University, Montreal, QC, Canada

The Chair's Perspective

Rationale (and CAHS Prospectus) for this Assessment

The majority of Canadians recognize that health research is one important mechanism by which significant improvements in disease treatment, evidence-based disease prevention, and health promotion are achieved, and most support the requirement for health research to be conducted. Investments in health research thus have increased significantly across Canada (and globally) during the past decade. Naturally, and justifiably, with greater investments have come increased expectations for returns. As noted by the prospectus for this report, these expectations have included:

- better health
- greater life expectancy
- translation of research findings into improvements in quality of life
- informed public policy on health-related issues across the full spectrum of government and private sector activity
- new commercial opportunities within and beyond Canadian borders
- increased attraction of the next generation to pursue careers in health research and the health sector
- a better "state of readiness" for the unexpected threats to health that inevitably develop in the contemporary world

Background

Given similar expectations on a global scale a number of countries have, over the past decade, launched investigations into how best to "capture returns" on their national investments in health research. These international studies were reviewed at the beginning of this assessment and provided an excellent starting point for our work. Another key background piece was generated by a process in Canada that began with the engagement of many stakeholders facilitated by CIHR in 2004 (refined by CIHR in 2005 and 2008) and resulted in both a synthesis and research evaluation framework. That process, combined with the on-going work on evaluation methods and frameworks by many funders across Canada (including many of our sponsors), created a strong background for this assessment.

The Scope and Process of this Assessment

With the help of individual panellists, our staff interviewed and synthesized the needs of the assessment's sponsors to ensure that we could make a genuine attempt to address them all. Our mandate from these interviews was clear: carefully define "return on investment," research this topic, and take it to the next level—a level whereby any stakeholder group in Canada can more easily access a set of useful references in this complex field, and more importantly—develop a "menu of options" for people to choose from, depending on their assessment needs. Sponsors wanted us to consider both quantitative and qualitative approaches, and to make this assessment "practical, robust, and objective" —supported by as much evidence as possible. They also wanted us to address gaps in the CIHR framework and to help clarify what is known as the "attribution issue of health research".

We then had face-to-face meetings of our full blue-ribbon panel of international experts, many of whom have led (or are leading) value assessment processes in their own countries. In addition to meeting the sponsor needs, the panel was charged with making cogent recommendations based on global research in the field (or to recommend research where gaps were identified) that would advance the ability of stakeholders to capture the impacts of health research in virtually any domain. They were specifically asked to propose a "best framework," with a "menu of metrics" by which return on investments in health research can be measured. Given the time frame involved, they were not asked to test the framework or validate the metrics proposed. That process will presumably happen via the sponsors in follow-up to this report. Panel members were also not asked to perform original research but instead to look for reasonable adaptations of existing models and methods.

It rapidly became apparent that it would be difficult to meet all of the expectations of every sponsor in this relatively brief exercise. It also became clear that the conflicting mandate of a "practical but robust" list of impact indicators and metrics would be very difficult. Nevertheless, the panel and our staff made every effort to achieve these objectives. We employed a variety of strategies.

In addition to panel meetings, to ensure both breadth and objectivity of the assessment, we commissioned six papers from national and international experts on topics that we felt would augment our own perspectives in areas of perceived need. These included: the public perspective on health research funding, ethics and evaluating health research, the meso-level metrics for impact, comparing international evaluation frameworks, clinical research, health services research, and population and public health research. We did not commission a paper on basic biomedical research as we felt that it was, based on size, background and international attention, the research area for which it was easiest to define methods and potential metrics of impact. We felt that these commissioned papers do add value to our assessment. We synthesized a large amount of material to inform the report that we have placed into appendices for those who are interested in more of the source materials for our assessment.

In closing, speaking on behalf of the panel, it is safe to say that we hope that the sponsors will find this report and its appendices interesting and its five recommendations useful. We enjoyed the academic challenge, learned a lot in the process, and enjoyed debating various aspects of this extremely complex topic. We hope that this report has a positive impact by helping sponsors assess their research programs and in helping interested stakeholders and researchers advance this important area.

Cy Frank MD

Chair

Acknowledgements

We thank CAHS for its financial support, guidance, and approach, with particular thanks to the outstanding CAHS executive: CAHS President Dr. Martin Schechter, CAHS Standing Committee on Assessments and its Chair Dr. Andreas Laupacis, and CAHS Past-President Dr. Paul Armstrong. We also thank Ms. Heather Good for her help with our processes. We also acknowledge and thank experts in various fields who contributed to our assessment by sharing their thoughts on research dissemination and impact with us. (Refer to Appendix H, p. A325). Finally, we thank the sponsors of this assessment, as this report would not have been possible without their support.

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 & 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

Table of Contents

Volume 1

List of Figures

List of Tables

List of Boxes

Volume 2

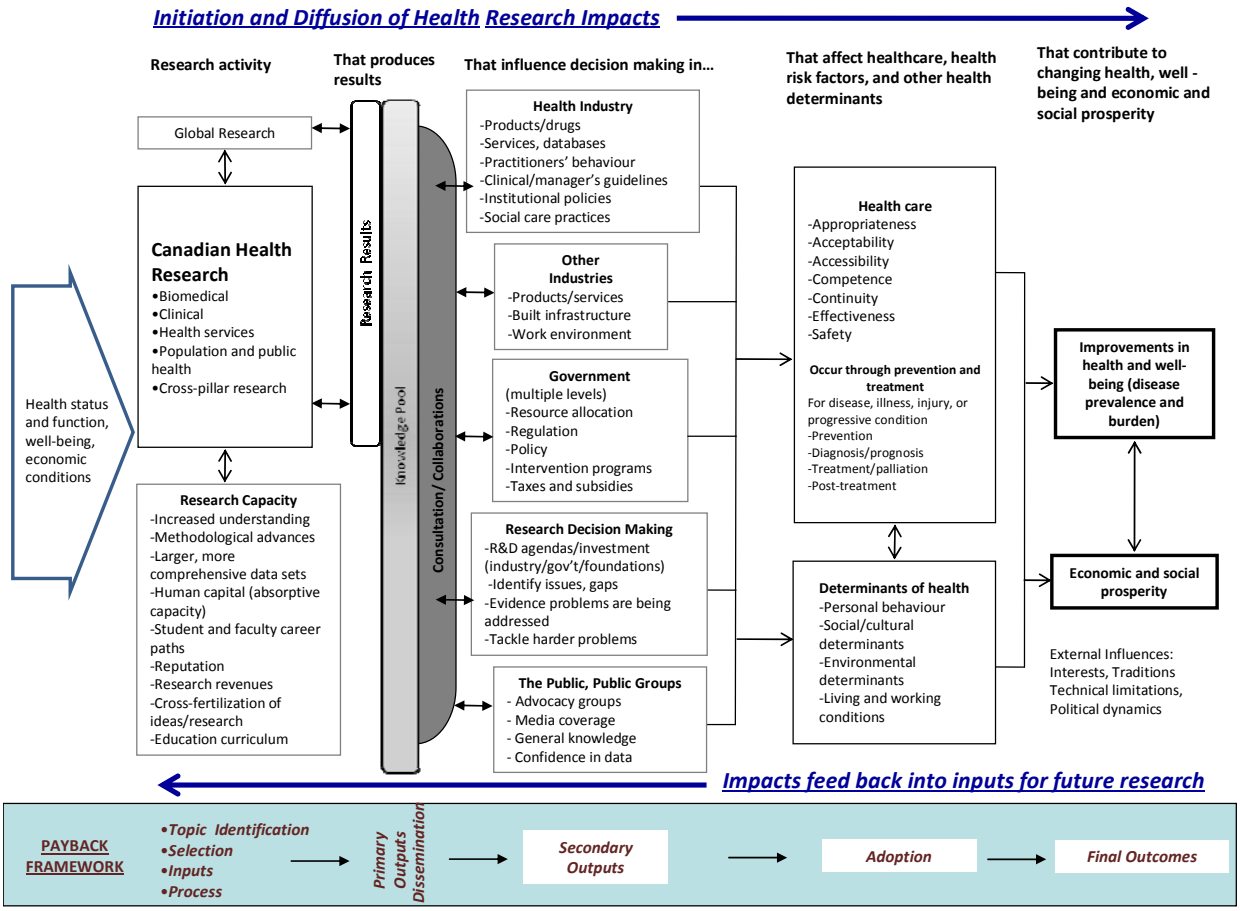
Appendices

The 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.

Executive Summary

Twenty-three different organizations sponsored this assessment. They all share an interest in defining the impacts of health research and learning how to improve the returns on investments in health research. Our remit from these sponsors was: Is there a "best way" (best method) to evaluate the impacts of health research in Canada, and are there "best metrics" that could be used to assess those impacts (or improve them)?

Based on our assessment, we propose a new impacts framework and a preferred menu of indicators and metrics that can be used for evaluating the returns on investment in health research.



The CAHS impact framework demonstrates how research activity informs decision making, eventually resulting in changes in health and economic and social prosperity (left to right arrow). The framework also shows how research impacts feed back upstream, potentially influencing the diffusion and impacts of other research, and creating inputs for future research (right to left arrow).

This framework builds on the combined logic model and impacts approach of the "payback model" (Buxton, M.J., and Hanney, S.R., 1996 – adapted by CIHR in Canada in 2005 and 2008), revised by our panel into a "systems approach" to capture impacts (this is shown at the bottom of the Figure). It is

designed to be used as a roadmap to track health-research impacts in five main categories: 1) advancing knowledge, 2) building capacity, 3) informing decision-making, 4) health impacts, and 5) broad socio-economic impacts.

Each of the main categories consists of subcategories, which identify and partition evaluation methods, and data that permit impact evaluation and identification of contributing factors. The first-level subcategories of each impact category are as follows:

- *Advancing knowledge indicators and metrics* include measures of research quality, activity, outreach and structure. We have also identified some aspirational indicators of knowledge impacts using data that are highly desirable but currently difficult to collect and/or analyze (such as an expanded relative-citation impact that covers a greater range of publications, including book-to-book citations and relative download-rates per publication compared to a discipline benchmark).
- *Research capacity-building indicators and metrics* fall into subgroups that represent personnel (including aspirational indicators for improving receptor and absorptive capacity), additional research-activity funding and infrastructure.
- *Informing decision-making indicators and metrics* represent the pathways from research to its outcomes in health, wealth and well-being. They fall into health-related decision-making (where health is broadly defined to include health care, public health, social care, and other health-related decisions such as environmental health); research decision-making (how future health research is directed); health-products industry decision-making; and general public decision-making. We also provide two aspirational indicators for this category (media citation analysis and citation in public policy documents).
- *Health-impact indicators and metrics* include those on health status, determinants of health and health-system changes, and they include quality of life as an important component of improved health. Determinants of health indicators can be further classified into three major subcategories: modifiable risk factors, environmental determinants, and modifiable social determinants.
- *Broad economic and social impacts* are classified into activity, commercialization, health benefit (specific costs of implementing research findings in the broad health system), well-being, and social-benefit indicators (socio-economic benefits).[2]

By choosing appropriate sets of indicators from the menu presented below, the CAHS impacts framework can be used to track impacts within any of Canada's four "pillars of health research" (basic biomedical, applied clinical, health services and systems, or population health) or within domains that cut across these pillars. It can also be used to describe impacts at various levels — individual,

[2] The indicators in these last two categories represent changes in impacts, which may or may not be due to research findings. Without tracking these changes, however, identifying the impacts of research would be impossible.

institutional, provincial, national or international — and to define funders' "returns" by (eventually) quantifying the value(s) of impacts to end-users as a function of dollars invested.

We believe that the framework's breadth, depth and flexibility fulfills our sponsors' request that the criteria be:

- useful to a full range of funders/research types,
- compatible with what is already in place in Canada,
- transferable to international comparisons, and
- able to identify the full spectrum of potential impacts.

To explain the logic behind the framework diagram, we must begin on the left, recognizing that health research occurs within the context of the Canadian population's health status, functioning, well-being and economic conditions. Canadian research is conducted within the context of a much larger volume of global health research. The Canadian context is further defined by the people and infrastructure that comprise current national research capacity. This influences selection of research questions within the four main pillars of Canadian health research, and domains that cut across these pillars. Moving to the right, the results of funded health research are first evident in published research documents, which then add to the global pool of knowledge. To varying degrees, the public and private industrial sectors (health and otherwise), as well as governments, draw upon this growing pool of knowledge to inform and influence future research agendas. New knowledge is also more broadly disseminated, usually by the media, to the general public and various interest groups.

It is important to note that these more direct uses of the products of health research are not their only impact. The end-products of health research manifest themselves partly in better and more appropriate interventions by the health system (both treatments and prevention), in health-improving changes to the broader determinants of health and, finally, in a more prosperous economy and society that is the indirect result of a healthier Canadian population.

While the CAHS panel clearly recognizes that making all the linkages from pieces of funded research at the left of the diagram to the outcomes at the right is a major challenge, with many gaps to be filled, it is precisely these outcomes that are the foundation for the public's support of health research. Moreover, we must emphasize that population health and economic prosperity are influenced by many factors other than health research. Thus a major question remains about how to parcel out the contributions of health-research to these bottom-line outcomes. This "attribution problem" for health research impacts is the main reason that the panel recommends that health-research funders support research into the basic science of health-research impacts.

To apply the framework appropriately, it should be populated by sets of both metrics and indicators of impact that are chosen to address specific evaluation questions. Users should take the following steps:

1. Define and prioritize *specific evaluation question(s).*
2. Use the framework to determine where to look for impacts.
3. Choose the impact categories (and subcategories) of interest: advancing knowledge, capacity building, informing decision-making, health impacts, and broad economic and social impacts. Be as specific as possible about where impacts are expected to occur and at what level (see subcategories above).

4. Choose a set of indicators and metrics carefully from the appropriate categories and subcategories of interest that will address the evaluation questions.
5. Ensure that sets of indicators meet the criteria of attractiveness and feasibility noted below.
6. Avoid inappropriate uses of the framework: Users should not ignore potential undesirable impacts of research by seeking only positive impacts (the so-called "Halo Effect"). They should also avoid "double-counting" of research impacts, and they should not ignore the "attribution issue" (in which contributions of other research and other factors known to impact health outcomes are ignored). Finally, users should consider "the counterfactual" — what would have happened if the research was not performed?. These issues are not easily resolved, and each requires additional research, including development of new indicators and metrics. Until that research is done, health-research impact evaluations should be based on balanced evidence and remain cautious about attribution. National and international collaboration in resolving these issues is highly recommended.

Our assessment also suggests that it is not possible to prescribe sets of indicators for different sponsoring groups (such as universities, government funders, voluntary organizations, etc). Users must define their own questions carefully and choose sets of metrics and indicators to address those questions in an objective way. In our report (on pages 106-108), we give three examples of how to use the framework to address theoretical evaluation questions for three potential users with quite different needs:

a. the biotechnology cluster (which is interested in the quality of its research, evidence of commercial gain, or whether employment opportunities were created);
b. a provincial funder of project grants (who is mainly interested in research quality, evidence of any health gains, and the ability to attract additional research investments to the province); and
c. a federal funder of research fellowships (who is interested in research quality, increased skill sets in Canada and dissemination of knowledge by its fellows).

These examples hint at the diversity of potential evaluation questions and why we cannot prescribe simple sets of metrics for all. Evaluation is expensive, and its costs must be reconciled by the potential magnitude of the findings' impact on users (funders and other stakeholders). Thus, we recommend collaboration in establishing evaluation questions of national interest in order to achieve economies of scale.

The second part of our mandate was to define appropriate metrics of health-research impact and to describe how to use them. We began this exercise by trying to define what precisely constitutes a 'metric', and by identifying how many such metrics exist. For our purposes, the term 'metric of impact' was taken to mean 'a numerical measure of impact'. Our assessment suggested that, while some metrics of impact (by that definition) clearly exist, many other useful non-numerical 'indicators of impact' have been described. The term, "indicators of impact", was therefore adopted to define all

indicators — some of which are numerical (metrics), and many of which are descriptive or qualitative. We have thus defined a menu of indicators that includes metrics as a subset.

Before listing the indicators, we will define how to use them. No single indicator of impact of health research — whether in any domain or in any impact category — is sufficient to demonstrate impact for any organization. Any single indicator can be dismissed as being partial, imperfect and likely to distort. For example, using publication counts as a single indicator is "perverse" and could have detrimental effects on scientists by sending the message that "volume equals quality." However — as part of a set of indicators that includes quality-of-research indicators, funding-received indicators (capacity building) and indicators for translating research findings for a variety of stakeholders — examining publication counts can help to identify those researchers producing quality *and* volume of research outputs. Indicator sets must also be appropriate to the right "level of aggregation" within each impact category. Some are appropriate at the individual researcher level, while others are not.

Selection of sets of indicators and metrics should be:

- focussed on the objectives of the organizations that will use them,
- appropriate for the stakeholders who are likely to use the information,
- balanced to cover all significant areas of work performed by an organization,
- robust enough to cope with organizational changes (such as staff changes),
- integrated into management processes, and
- cost-effective (balancing the benefits of the information against the costs of collection).

Indicators should also meet criteria of attractiveness and feasibility listed in the table below:

Table of preferred indicator characteristics - While it is not possible for every indicator to meet all of these criteria, it is important that each *set of indicators and metrics be selected to do so*.

Attractiveness:

Validity – does the indicator or metric reasonably reflect the underlying concept or construct that it is intended to measure?

Relevance – does the indicator or metric relate directly to a critical aspect of the research?

Behavioural impact – does the indicator or metric drive behaviour in a particular direction? Is it likely to result in any negative, unintended consequences? Does it create "perverse incentives?"

Transparency – is the methodology, and the strengths and weaknesses relating to the indicator or metric, readily apparent?

Coverage – does the indicator or metric cover a large proportion of output from research to be assessed?

Recency – do the data relate to current research performance, or look over a longer timescale?

Methodological soundness – is the calculation of the metric sound and statistically robust?

Replicability – can others reproduce the indicator or metric, and can it be used year on year in a comparable fashion?

Comparability – do other organizations collect comparable information or have targets to benchmark against?

Feasibility:

Data availability - do the data required to derive indicators or metrics exist, and do both the analysts and those being assessed have access to it?

Cost of data – how expensive is it to purchase the data outright or obtain on license?

Compliance costs – how labour intensive is it to extract/obtain the data?

Timeliness – can the data be obtained/provided relatively quickly?

Attribution – can the data be discretely ascribed to the unit being assessed? Direct attribution is ideal, but unlikely given current knowledge and methods; using attribution as a concept is important, as it provides a link between the impact seen and the research.

Avoids gamesmanship – does the indicator or metric provide scope for special interest groups or individuals to game the system?

Interpretation – can the data be open to misinterpretation or misuse by commentators and/or actors using the evaluation findings?

Well-defined – does the metric have a clear, unambiguous definition so that data will be collected consistently, and so that the measure is easy to understand and use?

Our menu of indicators and metrics is shown below, arranged according to the impact categories in the framework: advancing knowledge, capacity building, informing decision-making, health impacts, and broad economic and social impacts. These are the most wide-ranging indicators available, and they aim to capture as much information as possible across all health research. For each indicator and metric in the table, we have identified the "level" at which it is most appropriately applied (for example, individual researchers, research groups, institutions, etc.) and the CIHR pillar(s) of research to which each relates. Many are validated and practical, as they use existing data. However, not all of these indicators and metrics are currently available "off the shelf". In some categories we suggest indicators that represent important constructs, but whose characteristics have not yet been fully developed – we call these "aspirational indicators". It is necessary to prioritize and invest in developing and collecting appropriate data for all of these indicators, to prioritize aspirational indicators, and add to the list of indicators. It is also important to note that all but the most proximal indicators in the framework suffer from the issue of attribution (they are likely influenced by factors in addition to health research) and they require the most future development.

The quality of the indicator data is paramount. Special attention should be given to collecting data in a standardized fashion to allow comparisons of research impacts across funders within Canada and, ideally, outside of Canada. Standardizing the definitions for disease groups and types of research is the first step (such as the UKCRC classification (UK Clinical Research Collaboration 2007)).

Unfortunately, our assessment suggests that the data already being collected in Canada are not sufficient to provide all the information an evaluator might want, as there are likely significant gaps with regard to secondary outputs and outcomes. To address these gaps, there are two data collection techniques in use elsewhere: end-of-grant reporting, and common CVs for researchers. The standard format for end-of-grant reporting is a qualitative description of the research findings, process, and likely outcomes. Recent work in the U.K. incorporates a payback model-based survey that allows researchers to report on outputs and outcomes quickly at the end of grants, and again five years after the grant—thereby identifying late-arising impacts (Wooding 2008). It is also possible to expand the common CV (The Common CV System 2006) to include a variety of standardized outputs from researchers, such as presentations to public audiences, or consultations to government. Using these two approaches for collecting data ensures that there will be opportunities to identify research diffusion from researchers to a variety of stakeholders without being too onerous.

CAHS Menu of Preferred Indicators and Metrics of Impact - Appropriate indicators and metrics are arranged according to CAHS framework impact category. They should be selected in sets and mapped onto the CAHS framework to address different evaluation questions. They were selected for this menu from over 300 current indicators that were considered by the panelists. The subset of numerical indicators is called 'metrics' of impact. Note: this table contains our "starting menu" of preferred indicators and metrics - we recommend that they should be expanded over time.

ADVANCING KNOWLEDGE					
Category	**Indicator**	**Description**	**Level of Application**	**Comments**	**Pillars that indicators are relevant to**
QUALITY	Relative citation impact	* Average citations received by the unit being analyzed, compared to the world citation rate for the discipline(s) * World citation rates per discipline should be made widely available to interested parties	* Individual - not recommended * Group/department/grant - recommended * Institution/funding agency - recommended * National - recommended	* Must use discipline-specific benchmarks to account for different citation practices across disciplines * Only robust if based on a sufficient set of publications (individual researchers generally produce too few for robust analysis)	All pillars
	Highly cited publications	Individual publications are assessed against world citation thresholds to determine if they are in the top 1%, 10%, etc. of most highly cited publications in the world in that research area	Recommended at all levels	Must use discipline-specific benchmarks to account for different citation practices across disciplines	All pillars
	Publications in high-quality outlets (or desired outlets)	* Proportion of publications (publishers, conferences, journals) that appear in outlets judged to be of high quality * Could also include outlets that target specific stakeholders, such as those used by health practitioners	Recommended at all levels	Activity in a number of countries where disciplines are engaged in ranking the outlets of their discipline, including ranking publishers as well as journals (e.g. ESF Humanities project; Australia ERA journal and publisher rankings)	All pillars, but likely to be more important for pillars III and IV, where a smaller proportion of knowledge production is in journals
ACTIVITY	Share of publications	Number of publications from the unit under study as a proportion of a reference output (usually the level of aggregation above the unit under study)	* Individual - not recommended * Group/department/grant - recommended (share of institutional/funding body output) * Institution/funding agency - recommended (share of national output) * National - recommended (share of world publications)	* Normally done for field of research, rather than total publications * Can currently be easily calculated for indexed journal articles, but not for other types of publications	All pillars
	Publication counts	* Simple counting of outputs * Can be useful for new researchers who have no publication record allowing citation analysis	* Individual - recommended (number of publications by type: journal articles, books, book chapters, conferences, etc.) * Group/department/grant - not recommended * Institution/funding agency - not recommended * National - not recommended	* Counts by themselves are a poor indicator * The data are routinely collected in order to calculate other indicators (e.g. publication share, relative citation impact) * There needs to be a comparative aspect (e.g. is the level of output above or below that expected in that discipline)	This indicator is more important in pillar III and IV where a smaller proportion of knowledge production is in the journal literature. We strongly recommend that this indicator not be used as an indicator of quality in pillars I or II

ADVANCING KNOWLEDGE					
Category	**Indicator**	**Description**	**Level of Application**	**Comments**	**Pillars that indicators are relevant to**
OUTREACH	Co-author analysis	Determining the proportion of publications that are co-authored internationally, nationally, with industry, with other disciplines, etc.	Recommended at all levels	The selection of type of co-authorships to be analyzed will depend on the focus of the analysis	All pillars
	Field analysis of citations	Determining the proportion of citations that come from articles in the same field, and which other fields	Recommended at all levels	Gives an indication of the interdisciplinarity of the research by demonstrating the pick-up of research outside the core discipline	All pillars
CONTEXTUAL / STRUCTURAL	Relative activity index	* Determining the fields of research in which a unit is most strongly focussed * Uses the number of HCPs in each research area to show activity that is highest quality only	* Individual - not recommended * Group/department/grant - not recommended * Institution/funding agency - recommended * National - recommended	The benchmark for assessment will vary according to the research question, e.g. an institution may wish to compare its output to the national distribution, while at the national level the comparison might be to the world distribution or to similar countries	All pillars
ASPIRATIONAL INDICATORS	Expanded relative citation impact	Expanding citation analysis to cover a greater range of publications, including book-to-book citations	Aspirational at all levels except for the individual	There is work going on to try to improve the citation databases to include additional resources such as books, and this could be in place in the near future	Could prove especially important for pillars III and IV where a greater proportion of output is in the non-journal literature
	Relative download rate	Average number of downloads per publication compared to discipline benchmark	Aspirational at all levels except for the individual	* Ideally, downloads should differentiate between audiences, i.e. downloads from academic institutions, government agencies, general public, etc. * An equivalent indicator to highly cited publications for individuals could be "most downloaded"	Could prove especially important for pillars III and IV where a greater proportion of output is in the non-journal literature
	Research diffusion	Based on end-of-grant reports, which should include named individual researchers who should benefit from the research, and a sample of such individuals and their assessments of the actual usefulness of the research results, qualitative assessment of diffusion / uptake of research results	Aspirational at all levels	Requires thorough end-of-grant reports and follow-up	All pillars

CAPACITY BUILDING					
Category	**Indicator**	**Description**	**Level of Application**	**Comments**	**Pillars that indicators are relevant to**
PERSONNEL	Graduated research students in health- related subjects	* Numbers of graduated PhD/MSc/MDs, year on year * Should be able to disaggregate to subjects, gender, etc.	* Not recommended at the individual level * Can be used at institutional level * Most useful provincially or nationally	* As an aspiration we would also like to track the success of training programs in producing outstanding scientists and the progress that all research graduates make * Could be done in part using the Statistics Canada National Graduate Survey	All pillars
	Numbers of research and research- related staff in Canada	* Split into researchers, research assistants, and other staff * Can be disaggregated by province, research sector, etc.	* Not recommended at the individual level * Can be used at institutional level * Most useful provincially or nationally	Data already collected by Statistics Canada	All pillars
FUNDING	Levels of additional research funding	Funding from "external" sources that can be attributed to the capacity built in an organization, institution, or region. Could also include matched funding	Only recommended for funders, provinces, and nationally	Difficult to attribute to research funded by that province/organization, since researchers tend to be funded by multiple funding bodies (risks double counting)	All pillars
INFRASTRUCTURE	Infrastructure grants ($)	The amount in dollars of infrastructure funding pulled in by a research project, group, organization	Only recommended for institutions, organizations, provincially, and nationally	Captures the different aspects of infrastructure (kit, databases, buildings) since they all come from infrastructure grants, but misses out on infrastructure from other sources (e.g. university re-allocation of space, etc.) NOTE: This can be perverse if not aligned with operating money	All pillars
	% of activity grants with infrastructure support	Co-ordination of infrastructure grants with activity grants by identifying which activity grants have received additional infrastructure support to allow the research to occur	Only recommended for institutions, organizations, provincially, and nationally	* Does not account for research that has no new infrastructure costs or ones that are covered by universities * Data collection may be difficult and may have to be through surveying activity grant holders	All pillars
ASPIRATIONAL INDICATORS	Receptor capacity	Ability of those in policy and administrative positions to take research findings on board	Unlikely to be able to link to specific research findings, but could track the development of receptor capacity in Canada	There are surveys available to test receptor capacity, although these tend to be associated with specific training schemes	All pillars, particularly III and IV
	Absorptive capacity	Ability of researchers to take on other research from outside their organization, country, etc. and exploit that knowledge	Could address absorptive capacity for organizations, provinces, or nationally	Most commonly attributed through collaborations (particularly industry - academia collaborations) or R&D funding intensity	All pillars

INFORMING DECISION MAKING						
Category	**Subcat-egory**	**Indicator**	**Description**	**Level of Application**	**Comments**	**Pillars that indicators are relevant to**
HEALTH RELATED[3]	***Health care***	Use of research in guidelines	Analyzing citations to research in clinical and service guidelines	* Can be applied for individual researchers * More practical at aggregate levels (group/institution/ province/nation)	Allows identification of specific research informing health care and proportion of Canadian research informing health care	Mainly pillars I, II, and III
	Public health	Survey of public health policy makers	Asking public health policy makers what research has been used to inform their policies.	* Unlikely to be useful for individuals * May be useful for groups * Dependent on level of detail provided by policy makers	Surveying may be difficult unless policy makers are incentivized to take part	Likely to be pillars III and IV
	Social care	Use of research in guidelines	Analyzing citations to research in social care service guidelines	* Can be applied for individual researchers * More practical at aggregate levels (group/institution/ province/nation)	Allows identification of specific research informing social care and proportion of Canadian research informing social care	Likely to be pillars II, III, and IV
	Other	Researcher reported use of findings outside health	Example: health research findings could be picked up by transport or employment policy to improve safety or working conditions	* Could be applied to individuals * Better used at institution/funder levels	Since there are many different areas within "other," no single top-down indicator can collect all impacts, however, researchers may not know if their research is used outside their area of research	All pillars
	Health-related education	Research cited in ongoing health professional education material	Continuing health professional education materials produced cite research to support new practices	* Can be linked to individuals but likely to be small numbers * More appropriate at group/institution/funder levels * Recommended at provincial and national levels	* There may be issues accessing the references for these materials * Early health professional education covered in "research education"	All pillars
RESEARCH	***Research funding***	Citation analysis of successful funding applications	Identifying cited research in successful funding applications to identify underpinning research informing new research direction	* Can be used for groups and larger aggregations * Not recommended for individuals since number of citations is likely to be small	* Accessing references in successful applications can only be performed by research funders themselves * Data would have to be shared between funders	All pillars
	Research policy	Consulting to policy	Number of consultations to policy makers (from organizational to national policy) by researchers - year-on-year analysis	* Recommended for individuals; can help to identify which individuals are strongly linked into policy circles * Can be aggregated to groups above, but since there is no desired level of consultation is less useful at higher aggregations	* Needs to be addressed through surveying researchers * Top-down approach will miss "un-official" consultation	All pillars
		Requests for research to support policy	Number of requests for research for policy makers; primarily systematic reviews	* Only relevant at a provincial or national scale * Determines the level of interest in research, therefore not something research funders can influence directly	Can be addressed through official requests for research (systematic reviews commissioned) or through researchers' responses to requests	All pillars

[3] Within the four subcategories that represent the different aspects of a broad health system (health care, public health, social care, and other health related systems), there is a three-layer hierarchy of data sources for informing decision making metrics. The top level involves published evidence that identifies research; the middle level, surveying decision makers to identify what has influenced them; and the bottom level, asking researchers to report on how their research has informed decisions. The "most appropriate" indicators identified here are based on the most likely available information for each aspect of health-related decision making (so, if higher levels of information are not readily available, we recommend collecting information at the level below).

INFORMING DECISION MAKING						
Category	**Subcat-egory**	**Indicator**	**Description**	**Level of Application**	**Comments**	**Pillars that indicators are relevant to**
	Research Education	Research used in curricula for new researchers	Citation of research in textbooks and reading lists for university students in health-related disciplines	* Not recommended for individuals * Most useful at group/institution/funder/ province/national levels	Reliant on accessing lists of textbooks and papers used in teaching, as well as mining citation data from them	All pillars
HEALTH PRODUCTS INDUSTRY	***n/a***	Number of patents licensed	* Counts of licensed patents * Can be benchmarked against previous years or against internationally held patents	* Can be used for individuals * Most useful at group/institution/province and national level where sample sizes are larger	Data already maintained on patents licensed in Canada and reported on by Treasury Board	Likely to be pillars I and II
		Clustering/ co-location	Co-location analysis to show where industry is located in relation to academic centres	Only useful at provincial and national levels	Can provide an overview of where innovation and knowledge transfer is likely to occur	Likely to be pillars I and II
		Consulting to industry	Number of researchers consulted by industry; year-on-year values	* Can be used for individuals to identify those translating to industry * For group/institution/provincial levels can show environments conducive to knowledge translation (KT)	Data can be gathered through company reports or through researchers (as part of expanded CV or end of grant reporting)	Likely to be pillars I and II
		Collaboration with industry	Co-author analysis (bibliometric) of collaboration between industry and academia	* Not recommended for individuals (sample size too small) * Recommended for groups/institutions/ provinces/nationally	Reliant on industry publishing research findings in journals	Likely to be pillars I and II
		Use of research in stage reports by industry	Citation analysis of stage reports in development of products by industry	* Not recommended for individuals (sample size) * Recommended for groups/institutions/ provinces/nationally	Relies on accessing stage reports for industry (should be publicly accessible) and the ability to mine citations from them	Likely to be pillars I and II
GENERAL PUBLIC	***Advocacy groups***	Research cited in advocacy publications	Research mentions in publications (leaflets etc.) produced by advocacy groups, including patient organizations	* Not recommended for individuals (sample size) * Recommended for groups/institutions/ provinces/nationally	Misses other work for advocacy groups that is not cited, but consultations for advocacy can be captured in an expanded CV	All pillars
	Public education	Public lectures given	Number of lectures given to public audiences	Individual levels and above	Data could be collected through an expanded standard CV or through end-of-grant reporting	All pillars
ASPIRATIONAL INDICATORS	***Media***	Media citation analysis	Analyzing mentions of research in newspapers	Recommended at the individual level and aggregations above since media tends to mention individuals	* Potential international database of major national newspapers being developed * Requires individuals to identify research mentions in newspapers on a daily basis	All pillars
	Public policy use	Citations in public policy documents	Analyzing citations to research in public policy documents (grey literature)	* Could be applied at the individual level or above * More useful at the group level and above	The advent of Google Scholar as an analysis tool that can access citations in grey literature may help to analyze research informing policy decisions	All pillars

Indicators and metrics in the above sections of the table have a direct link to research. In the *health impacts* and *broad economic and social impacts* tables below, where links to research findings are

much harder to identify, we list the information that is most important to capture to identify changes in health, wealth, well-being, and social circumstances. It is necessary to perform additional studies to determine the link between research and the indicators below.

HEALTH IMPACTS						
Category	**Subcategory**	**Indicator**	**Description**	**Level of Application**	**Comments**	**Pillars that indicators are relevant to**
HEALTH STATUS	*Morbidity to include functional impacts*	Prevalence	Number of cases for a condition in a population (shown as a percentage)	Population level (from subgroups to full population)	Useful to show the impact of a condition on a population	Applicable to all pillars
		Incidence	Number of new cases for a condition per 100,000 population	Population level (from subgroups to full population)	Useful for identifying the new cases of a condition	Applicable to all pillars
	Mortality	PYLL	* Potential Years Life Lost * Number of years of life lost due to premature death (before 75)	Population level (from subgroups to full population)	Already collected across Canada through CIHI and Statistics Canada	Applicable to all pillars
	Quality-adjusted mortality	QALYs	* Quality-adjusted Life Years * Provides a value between 1 (perfect health) and 0 (death) of quality of life for each year lived after an intervention	Can be applied to specific interventions provided that data are collected, and can be used to describe populations	* Useful for linking to research impact since QALYs are linked to interventions (which can be more easily traced to research findings) * At the population level, data source is Canadian Community Health Survey	Applicable to all pillars
		PROMs	* Patient-reported Outcome Measures * Using a standardized questionnaire to determine patient views on quality of care and quality of life pre and post-treatment	Individual patients for clinical practice, but aggregations (e.g. hospital; disease state) for evaluation of research impacts	* Being developed to be more widely used in the UK NHS * Relies on patient reporting of their well-being	Applicable to all pillars
DETERMINANTS OF HEALTH	*Modifiable risk factors*	Example: obesity; alcohol consumption	Measures of prevalence of specific factors; e.g. for obesity, prevalence of BMI>30 for different population groups	* Can be at individual level * More useful for populations or sub-populations	Must be specific for the health problem under investigation	All pillars, but mainly pillar IV
	Social determinants	Example: education levels; social cohesion	Measures must be specific for the determinant; e.g. literacy levels for education	Needs to be by region (as aggregation could lose information)	Linking these social determinants to health research is difficult and requires additional research	All pillars, but mainly pillar IV
	Environmental determinants	Example: air pollution levels	Level of known toxic pollutants in the air (parts per million)	Needs to be by region (as aggregation could lose information)	Dependent on environmental risk factor under study	All pillars, but mainly pillar IV
DETERMINANTS OF HEALTH	*Acceptability*	Example: self-reported patient satisfaction	Surveying patients to identify their experience of the health service	* Could be applied from health care provider level to regional * Not useful beyond regional levels since information would be lost in aggregation	Some self-report surveys are not rigorous data collection tools and should be used with caution	Particularly pillar III
	Accessibility	Example: wait times	Wait times for specific conditions and/or interventions	Useful at provider, region, or population levels	Only applicable to secondary care	Particularly pillar III
		Example: appointment statistics	Time to appointments for different groupings (e.g. socio-economic, gender, ethnicity)	Useful at provider, region, or population levels	Potentially difficult to access disaggregated statistics from physicians	Particularly pillar III

HEALTH IMPACTS						
Category	**Subcateg-ory**	**Indicator**	**Description**	**Level of Application**	**Comments**	**Pillars that indicators are relevant to**
	Appropriateness	Example: adherence to clinical guidelines	Identifying whether practice conforms to the most up-to-date evidence base	* Can be used in audit for individuals * For evaluation it is most useful at provider, region, or national levels	Requires an audit of clinical practice, which needs to be based on a standardized survey	Particularly pillar III
	Competence	Example: civil law suits against the health system	Counts of civil law suits by clinical area over time	Could be used to show data from individuals upwards, depending upon the defendant involved in the suit	Civil law suits only identify the most extreme examples of incompetence, but measures of competence itself are difficult to come by	Particularly pillar III
	Continuity	Self-reported continuity of care	Surveying patients to identify their perception of the continuity of their care	Could be applied for individuals, health care providers, or regions	Self-reported data relies on standardized data collection across Canada	Particularly pillar III
	Effectiveness	Example: re-admission rates	Numbers of re-admissions by condition over a set time period; year-on-year change	Useful at provider, region, or population levels	Can only provide information on conditions that require secondary care	Particularly pillar III
	Efficiency	Actual vs. expected hospital stay	Length of stay for a patient compared to the expected stay for the condition	Not useful for individuals, only for provider, region, or national comparisons	Only provides data on secondary care and cannot take into account individual complications or co-morbidity	Particularly pillar III
		Cost input versus output	* Data on the inputs to health care services and on the different factors identified as outputs (e.g. available beds, emergency admissions, etc.) * Can be fed into a stochastic model to identify efficiency	Provider-level analysis only	Much of the data for any analysis is already collected for health care providers	Particularly pillar III
	Safety	Example: adverse drug effects	Numbers of adverse drug effects; year-on-year change	Provider, provincial, and federal levels	Adverse drug effects are an easily measurable safety issue, and one of the most visible	Particularly pillar III
		Example: hospital-acquired infections	Levels of HAI; year-on-year change	Provider, provincial, and federal levels	HAIs are a very current safety issue and are easy to measure and link to specific policies and research findings	Particularly pillar III

BROAD ECONOMIC AND SOCIAL IMPACTS					
Category	**Indicator**	**Description**	**Level of Application**	**Comments**	**Pillars that indicators are relevant to**
ACTIVITY IMPACTS	Economic rent (Labour rents)	The economic benefit (in $) of employing people in health research rather than in another capacity	* May be applicable at the funder or disease area level * Most useful at a provincial/national level	More comprehensive than simple employment benefits since it accounts for the counterfactual of what individuals would do if they weren't involved in research	All pillars
COMMERCIALI ZATION	Licensing returns ($)	Dollars spent on licensing patents held by Canadian organizations/individuals	* Not recommended for individuals * Recommended for groups/ institutions/ provinces/ nationally	Can be linked to specific research findings	All pillars, likely to be emphasis on pillars I and II

BROAD ECONOMIC AND SOCIAL IMPACTS					
Category	**Indicator**	**Description**	**Level of Application**	**Comments**	**Pillars that indicators are relevant to**
	Product sales revenues ($)	Sales revenues of products developed in Canada	* Recommended for provinces and nationally – could be used for specific funders * Not recommended for individuals; groups or institutions	Difficulty in linking to research findings means not useful for assessing research groups	All pillars, likely to be emphasis on pillars I and II
	Valuation of spin-out companies ($)	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)	* Recommended for provinces and nationally – could be used for specific funders * Not recommended for individuals; groups or institutions	Accessing valuation of new spin-outs may be difficult but are presumably available through venture capital firms that support the spin-out companies	All pillars, likely to be emphasis on pillars I and II
	Economic rent (Producer rent and spillover effects)	* Producer rent is the economic benefit to a company on top of expected revenues * Spillover effects are the external effects of investing in R&D on groups not invested in (e.g. investment from abroad in private R&D having benefits in Canada)	* Recommended for provinces and nationally – could be used for specific funders * Not recommended for individuals; groups or institutions	Calculating producer rent and spillovers has been performed for health R&D, but requires understanding of economic techniques underpinning analysis	All pillars, likely to be emphasis on pillars I and II
HEALTH BENEFIT	Health benefit in QALYs per health care dollar	Improvement in health measured through QALYs gained and divided by the cost of achieving that health gain	* Not recommended for individuals or groups * Useful for institutions/funders/ provinces/ nationally	QALYs can be monetized (controversial methodology) so a monetary net benefit could be compared to other uses of capital	All pillars
	Health benefit in PROMs per health care dollar	Improvement in health measured through PROMs gained and divided by the cost of achieving that health gain	* Not recommended for individuals or groups * Useful for institutions/funders/ provinces/ nationally	PROMs have not been monetized so this measure can only be compared to other PROMs measures	All pillars
WELL-BEING	Annual report of HRSDC	Human Resources and Social Development Canada (HSRDC) has multiple indicators of well-being that can be used to identify well-being	* National level only as difficult to attribute changes to research findings	* No links to research (health or otherwise) except through the "health" section of the well-being indicators, which are covered in the Health Impacts category * Data already collected and publicly accessible	All pillars, emphasis likely on pillar IV
	Happiness	As measured using established survey techniques for happiness-depression	* Recommended for provinces and nationally * Not recommended for individuals, groups, or institutions	* Self-report happiness scales used by Statistics Canada * Very difficult to make any link to health research findings currently	All pillars
	Level of social isolation	Loneliness scales for measuring social isolation of individuals	* Recommended for provinces and nationally * Not recommended for individuals, groups, or institutions	* Tools exist for measuring * Very difficult to make any link to health research findings currently	All pillars
SOCIAL BENEFITS	Socio-economic status	Identifying socio-economic status of individuals in Canada	* Recommended for provinces and nationally * Not recommended for individuals, groups, or institutions	* Causality of socio-economic status to health outcomes is well known * Not understood if health research can alter socio-economic status * Collected to identify if changes in socio-economic status correlate with research impacts	All pillars, emphasis likely on pillar IV

`

One key fact that has emerged from this assessment is that the science behind defining returns on investment in health research is embryonic and thus presents a significant opportunity for advancement. When we began this assessment, there was considerable interest in the topic and some base to build on, but there was no established model or validated method for tracing the impacts of health research. We think that this assessment has cast some light on the topic and on the issues that need to be resolved in order to truly define "returns on investment from health research". The commissioned papers that are appended to this report also contain valuable insights into the topic, including considerations for the development of other potential indicators.

Based on the path forward suggested by this assessment, our panel hopes that the "science of health research" will be advanced, in part, by the use of the impact framework and menu of indicators and metrics developed here, and by adding to the indicators over time. The longer term impact of using the framework and indicators will hopefully be the successful identification of validated metrics of impact, resolution of at least some of the attribution problem, quantification of the returns on investment in health research with both economic and value-based societal outcomes, and more rapid improvement of health research impacts over time.

Based on our assessment, the panel made five recommendations.

Panel Recommendations

1. **The framework and indicators identified by this assessment should be used by all funders of health research in Canada for evaluation of their health research impacts.** The measurement framework developed during this assessment incorporates the positive attributes of current best practices internationally (reviewed in Chapter 2) and reflects the needs of all Canadian and international funders of health research. It can be used to address existing and future evaluation questions for the purposes of accountability, advocacy, and for learning. The framework also provides a summary overview of the main sources, paths, effects, and impacts of individual or cumulative strands of research. It identifies the stages in the progression from funding to knowledge production, dissemination, uptake, translational modification, adoption, and impacts on outcomes over time, all of which are essential for a full appreciation of the impacts of funded health research.

2. ***Sets* of indicators and metrics chosen from our menu should be used by all funders of health research in Canada for evaluation of their health research impacts.** As reviewed in Chapter 4, it is critical for impact evaluators to recognize that multiple indicators and metrics are required for any evaluation, because any single indicator or metric can be dismissed as partial, imperfect, and likely to distort findings. Different subsets of indicators and metrics may need to be selected by evaluators to address different evaluation questions, since no single subset of indicators is ideal for every impact evaluation.

   It should also be noted that the "science of health research impact indicators and metrics" is still embryonic. The impact indicators we have assembled provide only a starting menu of indicators as of 2008. There are several areas where indicators are required, but the underlying methodologies have yet to be developed. In other areas where the methodology is clear, the underlying data are not routinely collected. In general, as the indicators and metrics move from more proximal impacts (for example, relating to published research results) to more distal impacts (for example, impacts on population health and the economy), the underlying data and methodology grow weaker and require more development.

3. **Canadian health research funders should begin collaborations immediately to advance the practical (methodological) prerequisites for measuring returns on investment in Canada.** There are a number of specific areas where evaluation of the impacts of health research is hampered by varying practices across research funders. These include the variable nomenclature used to describe different fields and topics in health research, data on the health research process itself, and inconsistent or limited accounting for the resources actually consumed in undertaking health research. As described in chapter 4, there is also a need to create and expand a central resource of scientifically validated indicators (with evidence that is easily accessible to the world). These needs drive a number of sub-recommendations for this new collaborative:

   a. *A commitment is required from all stakeholders and funders to standardize and refine methods, and to routinely collect high quality and appropriate data*. Current gaps in capturing research outputs and impacts must be addressed at the outset of research being performed, with good data collection methods and sharing strategies (with incentives to provide and share good data). This includes making a commitment to:

collect standardized information on common CVs, perform consistent end-of-grant reporting, use common sampling methodologies, standardize case study methodologies, establish common definitions and classifications of research disciplines, and standardize key words, etc. Research communities and stakeholders should be involved in defining what questions are to be answered, what data are required to answer them, what incentives should be considered and what data collection methods should be endorsed to build databases. Their time commitments to data collection obviously should be minimized by developing the most efficient methods possible. They should then participate in collecting good data.

b. *A library of impact indicators and metrics should be created, beginning with the starting menu developed here and then adding to it (using the criteria identified*). This library should be maintained on an easily accessible website to provide an on-going resource for sharing definitions and interpretations of the best indicators and metrics and their use. It should be updated regularly, because the number and sophistication of indicators and metrics will continue to evolve as new methods, new denominators, or new sources for indicators are developed (for example, the web as a data source). This investment will help to avoid unnecessary duplication and wasted efforts from using invalid or non-comparable indicators within Canada. It will also provide a readily accessible platform from which to contribute internationally to making health research impact evaluation more comparable and effective.

c. *A core set of key health research impact questions – based on what is practical and feasible – should be developed in Canada using our framework and indicators.* Canadian health research funders, together with their communities and stakeholders, must identify impact evaluation questions and define what indicator sets are required to answer questions for their own purposes (such as accountability, advocacy, learning or some combination of these) in their respective organizations. They must also implement appropriate data collection processes to support the construction of required databases.

i. *Strategic and ethically sound selection of indicator sets is required in order to avoid biasing future health research.* Because the selection by funders of evaluation criteria generally, and indicator sets specifically, has significant steering effects on researchers, funders and governments, “preferred sets of indicators and metrics” must be chosen carefully—particularly those with long-term implications (such as outputs, adoption and outcomes).

ii. *Evaluation questions and the choice of indicators and metrics to be used reflect political and social choices; the motives underpinning such choices should be made transparent.* Stakeholders need to understand the rationale for choices and be able to debate them.

4. **Canada should immediately initiate a national collaborative effort to begin to measure the impacts of Canadian health research.** Rather than waiting for perfect data collection methods and ideal, comprehensive, and validated indicator and metrics sets to be developed (see Recommendation #3), there is a need to immediately begin to use the framework and indicators to measure the impacts of existing Canadian investments in health research. This requires that

leaders from national organizations, industry and government organize a comprehensive effort aimed at a combination of learning, accountability and advocacy evaluation objectives. Research communities, other stakeholders and members of the public must also be engaged in that exercise. One option is that funders, led by one national organization, form a national council to lead strategic planning and execution of the framework, with a formal secretariat and commissioned data collectors to begin this work.

5. **Canadian health research funders should collaborate internationally to advance the "basic science of health research impacts."** Given the many unresolved issues of attribution, time-lags, the counterfactual and data collection limitations discussed in Chapter 3, it is recommended that Canada establish international collaborations to advance the "basic research" in this field. Such collaborations with international experts will help to:

    a. *resolve complexities:* (i) the diverse fields of health research generate a wide range of outputs that affect health, wealth, and well-being; (ii) advances in health are rarely (if ever) attributable to a specific funder or a specific discovery; (iii) the components of health research interact with one another; (iv) the uptake of health research is influenced by many complex factors including the incentives placed on potential adopters of research, the political environment, perceived and real inequities, and many other factors within the context into which research results are added.

    b. *improve the global system of health research impacts:* Understanding the world system of knowledge translation can help to improve it, by identifying key steps, eliminating bottlenecks, improving models to integrate across pillars, and testing other knowledge system innovations as they are discovered.

    Recommended plans include:

    i. *an international funding stream* – Collaborative international funding would be beneficial in advancing the elements identified in Recommendation #3.

    ii. *a research plan to identify "contribution indicators" and close "attribution gaps"* – The distal indicators of the impacts of health research are very broad, and many factors other than funded health research can and do affect these outcomes. It is fundamental that methodologies be developed to separate the contribution of health research from other causal factors.

    While it will never be possible to quantify all health research impacts with precision and accuracy, it is feasible that gaps can be closed by a concerted research effort to do so. (Caution: While gaps are being closed, all funders and practitioners must be aware of the issue of attribution.)

    iii. *a research plan to use the recommended framework for learning (impact improvement) purposes* – As noted above, the framework can be used to study where, when, how, and why knowledge may or may not be translated over time. A systematic approach to evaluating knowledge flows, barriers and facilitating factors that influence outputs and outcomes can be defined over time. Learning will help to guide better health-research investment decisions in the future.

1. Chapter 1: Background

1.1. Defining Health Research

In order to define a method to capture "returns on investment" in health research, first we must be clear about what the term "health research" means. This is surprisingly difficult, since "research that is relevant to health" can actually encompass a broad range of research fields and methodologies. Thus "the application of any research discipline to solving health issues" technically falls within the broad scope of the definition.

CIHR narrowed this definition somewhat by defining what have become known as "four pillars of health research" (Canadian Institutes of Health Research 2007a):

- basic biomedical research (investigating mechanisms of health or disease)
- applied clinical research (on, or for, patients)
- health services and policy research (investigating health services themselves)
- population and public health research (investigating populations and broader health determinants)

For the purposes of this assessment, this CIHR classification system was adopted and used to test the frameworks and indicators discussed in Chapters Two and Three.

While acknowledging that many other domains affect health, for the purposes of this assessment, we further restrict the idea of health research to be relevant to returns by organizations that would be recognized as "targeted solely at some aspect(s) of health." This might exclude research from other sectors that bear on health (for example, the environment, the economy, etc.), but is a necessary qualification for the framework and indicators defined below—designed for the funders of "health research" alone. We will also later define "returns on investment."

(Note: For a full glossary of definitions, please refer to Appendix F, p. A311).

1.1.a. The Canadian Health Research Landscape

Health research is an international effort, albeit one with leaders and followers, with publications produced the world over. In an analysis of world scientific publications spanning 1992 to 2001 (King 2004), 98% of the world's top 1% of cited publications came from only 31 countries, with the majority of those coming from the USA. Canada has recently been ranked sixth in terms of world publications produced, and seventh in terms of total citations (Sciencewatch.com 2008). Respectively, these data support the concept that Canada is a solid international contributor in terms of volumes and quality of research being produced.

Health research is a large and surprisingly complex system in Canada, with health research representing over one-fifth of the R&D funding for the country (Science, Innovation and Electronic Information Division 2008). There are six major funders of Canadian health research:

- federal funders
- provincial funders
- industry (business enterprise)
- the higher education sector
- the private not-for-profit (charity) sector
- foreign investment

These groups interact in complex ways. For example, funding for higher education institutions comes largely from provincial government funds, while industry and the federal government have joint funding streams. Also, funding for research can take many forms, primarily activity and human resources funding from research-funding organizations like CIHR, but also infrastructure funding from other sources, such as the Canadian Foundation for Innovation (CFI) and the Indirect Costs of Research Program (ICRP). Figure 1 shows some of the interaction between the major funders of the Canadian health research system and the major bodies that provide funds to researchers and organizations. It does not show all funders, but represents the major funders or those that illustrate specific types of funding.

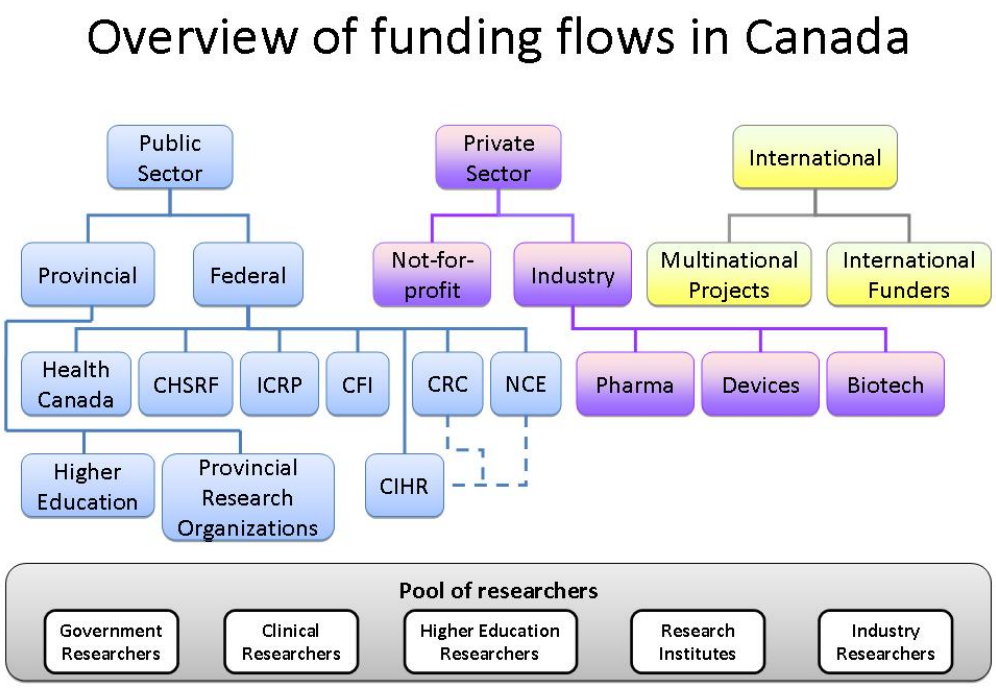


Figure 1. Major funders and funding flows in the Canadian health research system (Nason 2008)

In the public funding sector, there are federally funded bodies for activity (Health Canada, CIHR), bodies funded through federal endowment funds (Canadian Health Services Research Foundation – CHSRF), bodies for infrastructure (ICRP, CFI), and research schemes that are not exclusively for health research (Networks of Centres of Excellence – NCE; Canadian Research Chairs – CRC). For NCE and CRC, funding is through the three federal research councils, with CIHR providing funding for the health aspects of each scheme (Nason 2008). The bottom of Figure 1 shows the groups that perform research in Canada. Researchers can be funded by any of the funding groups, although federal researchers are far less likely to receive extra-mural funding.

Funding is not split evenly between these types of research funders; neither is research performed equitably across the identified sectors of researchers. In 2007, the major funders of research were the business, higher education and federal sectors, but the vast majority of health research in Canada was performed by the higher education sector (Figure 2).

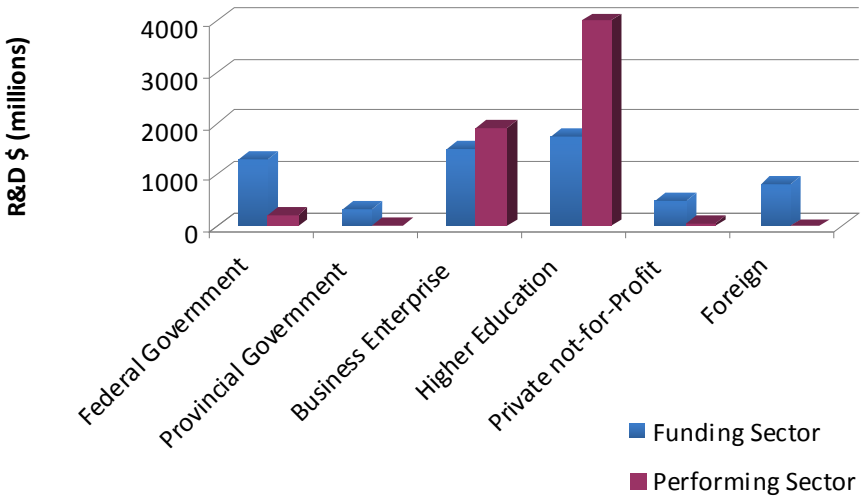


Figure 2. Funding and performance of health research in Canada in 2007 (Science, Innovation and Electronic Information Division 2008)

The diversity of sponsors for this assessment[4] highlights the need of all health research funders to understand their impacts. (Appendix C identifies how selected of funders in each category currently evaluate their performance. See pp. A235-A246)

With multiple funders of health research in Canada, it is not surprising that there are multiple missions for the funding allocated. For example, federal funders the Canadian Foundation for Innovation (CFI) and the Canadian Institutes for Health Research (CIHR) fund infrastructure and activity, respectively,

[4] See page 10 for a full list of sponsors for this project.

but focus on different impacts for their missions. CFI has a mission to improve the ability to perform research, while CIHR aims to improve health. Although not mutually exclusive, having different missions means that researchers have no assurances that they can get aligned funding for both the activity and the infrastructure costs of their research. Understanding the impacts (both desired and unexpected) will help to identify where links between funding bodies could be improved, and what impacts would be facilitated by such improvements.

Since identifying and addressing any discontinuities between funders is a potential impact of evaluation, it is also possible that effective identification of impacts could affect individual funders. For example, for a funder across all four pillars of health research, evaluating impacts may show that particular pillars should receive more or less funding to achieve specific aims. This is similar to findings in an evaluation of the U.K. Arthritis Research Campaign's portfolio, which identified the different types of impact that arise from programs, projects, and fellowships (Wooding, Hanney, et al. 2004). This is the first step in producing the base of evidence for effective funding to address specific health problems and to gain specific desired outcomes.

Appendix B provides a more detailed analysis of the Canadian health research landscape (p. A212). Evaluating this very complex system requires a high-level approach that can capture diverse stakeholders and their interactions (for example, joint funding initiatives between CIHR and industry funders: Canadian Institutes of Health Research 2008a).

1.1.b. What Returns are Expected from Canadian Health Research?

Better health and related improvements in longevity and quality of life for people are the undeniable goals of all health research, and serves as the major motivation for those performing it in Canada. Successfully achieving that goal (or even a fraction of it), with evidence of meaningful, measurable, and attributable impacts on health outcomes in the Canadian context, would serve as the most compelling rationale for Canada to invest in health research. Beyond direct impacts on the health of the country's population, however, there are many other potential outputs and outcomes from health research investments that, if quantified, would also demonstrate value in health and related life sciences and technology. These include:

- evidence that health research enhances the entrepreneurial advantage of Canada
- world-class excellence in knowledge creation and training
- recruitment or retention of exceptional scientists in Canada
- the commercialization of Canadian research findings (for example, licensing patents)
- the attraction of international health-related industries to locate in Canada[5], with health-related and economic impacts (for example, employment, wealth creation, etc)

Process-related indicators might include evidence of being an international model for setting appropriate health research priorities, establishing effective research partnerships, and demonstrating

[5] In Ireland, for example, the government deliberately invested in health research to encourage pharmaceutical manufacturing already in place in the country to remain there and to invest in R&D as well as manufacture (O'Neill 2007).

research accountability. These potential indicators link to the priorities identified in Canada's Science and Technology Strategy (Government of Canada 2007a).

Evidence from more indirect outputs might also show value from investments in health research. For example, health research might benefit science, policy-making, and health services in Canada by showing evidence of an improved, science-related "absorptive capacity" (making the uptake of research findings from other countries easier) (Zahra and George 2002), or enhancing "receptor capacity" (the ability to use research findings in decision making, such as public policy). It might also simply demonstrate evidence of training highly qualified personnel who add to the pool of science experts, and thus enhance Canadian capacity to absorb and translate international advances in health into its own population. Impacts that are even more indirect, such as evidence that health research potentially improves the decision making of providers or patients in Canada, or retains critical health care staff to the Canadian health care system could also be considered indicators of success. It is easy to see that some of these impacts would be much more difficult to measure than others, and some would be much less likely than others to provide compelling reasons to make investments in health research. Some have direct impacts, while others might have great difficulty demonstrating cause and effect relationships (attribution) because of being intertwined with so many social, cultural, and economic factors.

Regardless of how complex they may be to measure, all of the above are potential "indicators of success" of health research to those who fund it at a national level in Canada. Evidence of any or all of these outputs or outcomes, normalized to Canada's health research investment and benchmarked internationally, could establish Canada's role as a leader in demonstrating impact in the world of health research. The bigger the impact (with evidence), the better to justify expenditures in health research and thus, any of the above could be used as "indicators of success" for health research.

Unfortunately, defining "measures of success" is even more complicated than that. Interviews during this assessment revealed that there are many different views about the potential value and "preferred indicators of success" of health research, even among Canadian funders of health research, let alone the many other stakeholders who are influenced by or benefit from the health research industry in Canada. An overview of some of these divergent views regarding the best approaches to measuring returns and "the most important evaluation questions" will help to set the stage for a discussion of how to determine an overall "best approach" and the "best metrics" (or for our purposes, the most appropriate indicators) to capture the returns on investment in health research in the Canadian context (our remit).

1.1.c. Many Stakeholders, Many Views

Interviews with sponsors and stakeholders in this assessment revealed that they have a broad range of potential health research impact evaluation questions, making it difficult to produce a "simple evaluation method" that will work for all parties.

Data suggest that perhaps the most important stakeholder in this assessment, the so-called "average Canadian," is already very positive about the potential benefits of health research. As noted by A. Picard in his paper commissioned for this report entitled, "Translating Science Into Hope: The Public Perspective on Health Research Funding," (Appendix A, p. A138) Picard says: "A survey commissioned by Research Canada showed that 85% of respondents believe governments should spend more on health and medical research," and "69% of those polled said they would be willing to spend out-of-

pocket to support the research endeavor;" albeit "a modest $1 per-week." Later in his paper, Picard notes an even greater qualification that, depending on how evaluation questions are asked, and if some of the real investment numbers on a national scale are made known to them, the same member of the public may have quite a different perspective. He states: "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." This implies that members of the public want to know what they are getting for their money and "how much is enough?"

Other sectors touched by health research currently view the Canadian health research enterprise in a similar way; positively—but with qualifications. Disease-specific voluntary organizations in Canada support the view that a proportion of philanthropic funds should be dedicated to Canadian health research, since it certainly increases the awareness that health research is needed, and because strategic investments in research, they believe, can increase the likelihood of finding meaningful improvements in health for Canadians over time. Research invokes hope. As positive as most voluntary organizations are, however, they express some concerns about not being able to easily demonstrate evidence of impacts from their investments in Canadian health research thus far. "Are they getting value for their money?"

Many industry and biotechnology leaders recognize that health research is an essential component of their product development, with a clear deliverable in most cases being a defined product or service that can be sold. The economic value created is a clear measure of their product effectiveness, and their profits reflect their corporate efficiency and business success, providing outcomes that are clear and measurable. They wonder "how much to invest and in what stages of public R&D?", as successful products are few and far between.

> *"If you think research is expensive, you should try disease."*
>
> Mary Lasker: Founder of the Lasker Foundation (Funding First 2000)
>
> Figures on expenditure for health research and health care in Canada can be used by research advocates to support that statement, with an estimated expenditure of $160B in 2007 on the health care system (Canadian Institute for Health Information 2007) and $6.3B in 2007 on the research system (Science, Innovation and Electronic Information Division 2008). These data imply that more research funding will decrease the burden of disease. Will it, and if so, how quickly?

Health researchers, while excited about the potential of their work to have impact, actually have the most conservative view of all of the sectors interviewed about the products of their work. Interestingly, they are (based on evidence and experience, in most cases) appropriately cautious about the likelihood of any health progress or impact as a result of their work, noting instead that impacts will likely happen slowly as their discoveries become translated by others into meaningful advances over time. They also require evidence to support their views—the essence of why they are scientists.

Some recognize that there should be a science to all of this: the science of demonstrating and understanding (to improve) health research impacts.

1.1.d. Different Evaluation Purposes

Adding to the complexity of having the many stakeholders with many health research evaluation questions noted above is the underlying fact that even among the relatively small number of health research funders in Canada, evaluation needs are different. Interviews supported international experience that health research funding evaluations are generally performed by funders for one (or more) of three different reasons:

- evaluation for accountability purposes
- evaluation for advocacy purposes
- evaluation for learning purposes

Each of these reasons implies different organizational goals, and likely requires different evaluation strategies.

1.1.d.i. Evaluation for Accountability

Government funding has a direct responsibility to the tax-payer, charity funding to the donor, and industry to the shareholders. To show that money has been spent wisely, organizations must evaluate the outcomes of their funding in relation to their anticipated goals. This requirement to evaluate "on mission" is a key driver of the recent shift toward evaluating public research funds.

Accountability is considered by the OECD to be an obligation to conduct work in accordance with rules and standards, and to report accurately on performance; it relates specifically to a demonstration of work conducted in alliance with "contract terms" (Development Assistance Committee Working Party on Aid Evaluation 2002). This highlights the link with the mission of research funding organizations. In fact, for government funding organizations, targets are often set by government for the organization to try to attain (Government of Canada 2007b). Accountability evaluation can use these set goals and what the organization has achieved in relation to them. Evaluation for accountability also has strong links to value-for-money issues, with accountability for funding often relying heavily upon the ability to show that the money was invested correctly.

Examples of evaluation questions that are linked to accountability include:

- Has the project fulfilled its aims?
- Has the funding been distributed to projects that have been successful?
- Is the research funding being distributed appropriately, or are there other ways to distribute funding for "better" results?

Asking these sorts of questions automatically implies specific measures that will be useful in evaluating the funding, however, there is an inherent difficulty in showing that the use of funding was more appropriate than any other use (the counter-factual proble

1.1.d.ii. Evaluation for Advocacy

This concept differs from accountability evaluation in that where accountability relies upon delivering upon promises, advocacy relies on the promises it may be possible to deliver. In a sense, the purpose of evaluation for advocacy is to increase awareness of the great things a research funding organization can achieve.

For example, where a charity performing an evaluation to show its accountability to donors would identify the full extent of funding, research projects, and impacts, a charity performing an evaluation for advocacy needs to identify the research that best highlights the future possibilities of the organization and the "best" impact of its research investment. Identifying the "best" impact is subjective, and varies depending on the group to which the charity (or any funder) is advocating.

Examples of evaluation questions that are linked to advocacy include:

- What has been the biggest impact of our research?
- Where is our strongest research area to showcase?
- What are our "good news" stories?

These questions differ considerably from those for accountability, and require different data in order to be answered. Where evaluation for accountability suffers particularly from a lack of counterfactual, advocacy evaluation need only identify what the funding has done well in order to convince funders that their investment has paid off.

1.1.d.iii. Evaluation for Learning

The first two reasons for evaluation are outward looking, facing funders, potential funders, and stakeholders. Evaluation for learning is inward looking, trying to identify how a funder is performing in achieving its mission, understanding the processes it has in place, and identifying where opportunities, challenges, and successes arise for their research. In its most basic terms, evaluation for learning asks, "What are we doing well? What are we doing badly? And how can we improve?"

For a research funder, this means understanding the research process and the paths to impacts in a more nuanced way than required for an accountability or advocacy evaluation.

Examples of evaluation questions that are linked to learning include:

- What have our different funding streams achieved?
- How have we as a funding agency helped to foster positive research impacts?
- Where could we improve our research portfolio to fill a research gap?

Evaluation for learning suffers specifically from the attribution of impacts to specific funding, since the evaluation is an attempt to understand how an organization/funder has affected the impacts arising from health research. However, it is a valuable tool for any funder that wishes to improve its impact, effectiveness, and efficiency over time.

Learning as an outcome of all evaluation methods is valuable, especially with regard to making decisions about where to put future research funding, since an evidence base on the effectiveness of past funding is available. For example, projects, fellowships, institutes, and program funding were

compared for the types of impacts that they had in the evaluation of Arthritis Research Campaign (arc) funding in the U.K. (Wooding, Hanney, et al. 2004). This work found that fellowship and program funding tended to produce knowledge, capacity building, and decision-making impacts, but not health, economic, or social benefits. It also found that projects and institutes tended to have benefits in all areas (Wooding, Hanney, et al. 2004). However, since this study used only 16 case studies of arthritis research, it should not be generalized to all research without further understanding different types of research and more data.

Figure 3. Links between the three reasons for evaluating research

1.1.d.iv. Comprehensive Evaluation

Not surprisingly, it is possible to combine all three evaluation purposes into one grand evaluation. With this method, the evaluation seeks to provide answers to questions from all of accountability, advocacy, and learning evaluations. This is the most resource-intensive evaluation method, as answering these questions often uses different data, data collection methods, and analyses. Current international health research evaluation frameworks do not address this comprehensive evaluation, because data collected to answer one set of evaluation questions may not be suitable to answer other sets (RAND Europe Report on International Evaluation Frameworks – Appendix A, p. A147).

1.1.e. Why Measure Returns on Investment for Health Research in Canada?

Evaluations of the impacts of health research are being performed all around the world (Buxton, Hanney, et al. 2004). Why then does Canada need its own approach? This question can be partly answered by the fact that evaluation relies heavily upon the mission of the funder and the questions that can be asked of the funding situation. Given the complex funding landscape in Canada, it is necessary to have a customized set of indicators for the different purposes mentioned above.

The real issues, of course, are: How many of the goals and possibilities for the outputs and outcomes of health research are actually being achieved in Canada? Is there evidence of achievement or progress in those domains? Further, do all Canadians have these same basic questions about the impacts "health research," or are there a large number of questions to be answered? Can one simple method

and a few basic evaluation questions be used to satisfy everyone? Is there an easy way to define Canada's achievement of its social and economic goals? Is Canada unique in health research, and if returns are going to be measured, can the method take this uniqueness into consideration?

These are difficult questions, and the difficulty is compounded by considerable evidence that Canada *is* unique and *does*, as we will review below, have a unique approach to health research. Despite being a relatively young nation, Canada's rich history in contributing to the world body of knowledge in health sciences, and its evolution during the establishment of CIHR, has made Canada a world leader in embracing vastly diverse perspectives in its health research efforts. This uniqueness also, however, drives the need for a unique evaluation strategy for its full health research spectrum.

1.1.f. What Evaluation of "Returns" is Already Taking Place in Canada?

As of 2008, evidence shows that Canadian health research funders have a number of different models and processes in place for evaluating their health research investments. These vary from ad hoc evaluations of specific funding by case studies, to the use of specific evaluation frameworks across the entire research funding enterprise. By identifying the existing strategies, it is possible to determine which aspects of Canadian health research evaluation are already well covered, and where there are opportunities for improvement.

As one of the leaders in Canada, and as the major federal funder of health research, CIHR has incorporated a modified version of the internationally recognized "payback framework" into its evaluation procedures (Canadian Institutes of Health Research 2008b, Canadian Institutes of Health Research 2005b). That framework, developed by Buxton, Hanney, and co-workers, is built on a logic model of how new knowledge flows across many domains. It is explained in more detail in Chapter 2. The payback model has also been adopted by other research funders in Canada. The Saskatchewan Health Research Foundation (SHRF) incorporated the economic benefits identified through the payback model into its evaluations (Peach and Marshall 2008), and the Alberta Heritage Foundation for Medical Research were the first in Canada to use the payback model to investigate its outcomes in the late 1990s (Buxton and Schneider 1999).

Variants of the two main strategies within the payback model, logic modeling and impact categorization, are used separately by a number of organizations in Canada. The Canadian Health Services Research Foundation has an evaluation system built on logic models of the way the organization works (Canadian Health Services Research Foundation 2005). The Fonds de la recherché en santé de Québec (FRSQ) has also produced a logic model that shows the impact chain of research—a model that has also been taken up by the National Alliance of Provincial Health Research Organizations (NAPHRO) for all provincial organizations to use (Beaudet 2007). The Heart and Stroke Foundation of Ontario (HSFO) has used logic modeling to address specific programs of research, such as the "Healthy Weights Area for Investment in Mission (HW AIM) Initiative" (Heart and Stroke Foundation of Ontario 2007). The balanced scorecard has been used in some health research evaluations, with the caveat that it has not generally been used to provide a balance across research, but rather that research has been one aspect of the scorecard for the organization (University Health Network n.d.).

We indirectly analyzed the perceived weaknesses in these current Canadian frameworks. In addition to the evidence for any "best metrics" identified through existing evaluation systems, evidence can be gathered by analyzing the weaknesses in current evaluation systems by asking funders themselves what they would look for in the findings from an assessment of health research evaluations (Chapter

2). The findings from this consultation indicate that improvements are needed in three main areas: improved framework development, improved indicator development, and better understanding of the issues that affect evaluations (such as attribution and time-lags: Chapter 3).

1.1.f.i. Canada's R&D Productivity Versus its R&D Funding

Canadian research contributes a small but important portion of the research effort globally, accounting for 5.3% of all scientific citations in the period 1997–2001, ranking sixth in the world (King 2004). Canada performed better on health research alone, however, especially in pre-clinical medicine and health research, where it ranked fourth in the world (narrowly behind Germany and the U.K., and considerably behind the USA).

Most comparisons of international research suffer from the problem of generalizing R&D, making it difficult to identify how well Canada fares in relation to other countries. Funding-wise, comparisons of research funding within countries is often performed using gross expenditure on R&D (GERD), and business expenditure on R&D (BERD), with more specific health comparisons made using 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). These are collected by the OECD and used by government to monitor funding inputs (Treasury Board of Canada Secretariat 2007), by research organizations to show the position of Canada in R&D (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).

It is useful to be able to identify where Canada lies in relation to other research-funding countries, since it allows a benchmark for the potential impact of Canadian funding compared with funding in other countries. Evaluating the impacts of funding at the more relevant levels of outputs and outcomes, however, requires an evaluation method that can provide comparable data on products and outputs from the Canadian health research system versus those from other countries.

1.2. Are "Health Research Impacts" Already Defined Elsewhere?

Many countries have launched similar assessments of their "returns from health research" within the past few years. The results are, on average, contradictory, and the evidence of health research impact is either highly qualified or mixed. Different approaches have been used to evaluate returns and different conclusions reached.

For example, some research has shown very large financial returns on investment from health research (Funding First 2000, Access Economics Australia Economic Consulting 2003), implicitly supporting the position that increased spending on health research will have positive impacts in perpetuity, and implying that the ultimate impact of investments in health research will be to eliminate serious diseases, improve population health, and decrease spending on health care. Others have endorsed more research funding up to a point, but have advocated a limit on public investments in health research based on the law of diminishing returns (that is, funding will eventually exceed the capacity for its appropriate use and people cannot be made more than 100% healthy). These groups recommend an evidence-based limit on public funding (with a method to be determined) to allow

required public funding for other research priorities in addition to health (such as climate change at this point in time).[6]

In addition to these unresolved questions regarding how much *total funding* is required (public + private), the question of how much *public funding* is "appropriate" has been even more contentious in the literature. On one hand, it is suggested that public funding is essential, and on the other hand, that public funding is counter-productive and crowding out private innovation (Organisation for Economic Co-operation and Development 2003). Most recent work suggests that the "crowding out" phenomenon does not occur extensively in health research, but it may have to be considered (for evidence see: Cumbers and Birch 2006, Congressional Budget Office 2006, Rosenberg 2002, Joint Economic Committee 2000). This supports the notion that both public and private R&D investment plays a role in furthering health research, but it does not answer the questions: How much of each is "enough"? Will a lot more public funding of health research help or not? Would partnership programs do even better? Have economic analyses shed any additional light on these topics?

1.3. Economic Evaluations to Date

As reviewed by Shiel and Di Ruggiero in a paper commissioned for this report, titled "Assessing the Return on Canada's Investment in Population and Public Health Research: Methods and Metrics," (Appendix A, p. A43) there have been two approaches to answering some of the economic questions posed above. The simpler "return on investment" approach provides a gross "indication of how the momentary value of the health gains that follow from successful implementation of research findings compares with its costs." The "payback approach" provides "a more nuanced and comprehensive description of the impacts of research funding." In their synthesis of return on investment approaches, Shiel and Di Ruggiero review a series of papers commissioned by the Lasker Foundation in the USA involving the work of several economists: first, to quantify the health improvements 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 (McClellan, Heidenrich 2003). Cutler and Kadiyala (1999) suggested that at least one-third, and perhaps as much as one-half, of the gains in quality of life and life expectancy since 1970 in the USA could be attributed to health research. Shiel and Di Ruggerio then noted that, based on the models and assumptions of the values of life and attribution used, the conclusion that health research paid huge dividends was reached. They summarized the key finding that "if one-third of the improvement ($1.5 trillion per year) was due to health research, then the social value of this improvement in health ($500 billion per year) would cover the costs of research 20 times over. A similar method was subsequently adapted and used to value the returns on investment to all research in Australia…Every dollar invested in research yielded net benefits to society (that is, benefits over and above costs) of at least $1.17 and perhaps as much as $1.40."

Their critical review of this evidence, plus their support for the proposed use of the "payback approach" in the Canadian context (the latter to be discussed in much more detail in the assessment

[6] It is worth noting that for private companies, there often is a "right" amount to spend on R&D, with start-ups requiring more funding aimed at research to secure a position in the market (Barber and Crelinsten 2004). It is not so simple to generalize about public funding, however, since it involves other pressures that do not exist in private organizations.

below), is noteworthy and is recommended reading as further background at this point. (Appendix A, p. A43).

1.4. Summary of the Landscape

The Canadian health research scene is unique, complex, and dynamic. It has many stakeholders, each with a large number of different evaluation questions based on stakeholder-specific needs. Funders have a variety of needs, including the need to see evidence of impacts from their significant investments in health research in Canada. Many have realized that this is not a straightforward issue and want more background on the topic, as well as a logical approach for sorting it out.

While there are a number of excellent frameworks and a few international reviews of "ROI" in other countries, there is no accepted international standard framework or indicators, and there is no agreement on a standard approach to determining the value of health research. Some interesting indirect methods have suggested that health research is a bargain with major economic returns on investment (Cutler, Kadiyala 1999), but most more-direct investigations are far less conclusive.

There are many reasons for this inability to reach definitive conclusions, including: variable definitions, lack of consistency in approaches, incomplete data (collected for other purposes), lack of prospective evaluation methods aimed at answering specific evaluation questions, and a host of other measurement issues.

Almost all recent investigators in this field have concluded that this topic requires further definition and study. Interviews with international leaders during this assessment confirm this view and note that most experts internationally are asking similar questions: What *are* the "returns" on investment in health research in any country (or collectively)? Where and when can they be expected? How can returns be identified and traced over time? Who has the best methods for tracing health research returns, and is any one country demonstrating returns better than any other? How does (my country) fare in contributions? How much is enough? Is there a proportional return on money invested? How do we know? Is the system of health-related knowledge production and translation optimized? Is there any evidence that it is or isn't? Is there a single best way to answer all of these questions?

There is a strong need to review the entire topic and to make recommendations that will clarify the issues and define an approach going forward.

2. Chapter 2: Frameworks

2.1. Rationale for a Framework to Understand Health Research

Basic biomedical research, clinical research, and health services and public health research collectively generate a wide range of outputs that affect health, wealth, and well-being. The uptake of health research is further influenced by many factors, including the incentives placed on the potential adopters of research. Unfortunately, demonstrating the causal pathways that may lead to impacts from the diverse fields of health research is thus extremely difficult.

This problem of complexity creates the need for a standardized solution, in which meaningful "pieces" of the entire "system of health research" can be identified and classified in way that captures both production and use of knowledge and the impacts of that knowledge over time. Classifying these elements and mapping their relationships within as an "evaluation framework" aims to identify the sources of knowledge, and the categories of impacts, and allows identification of best indicators within each of those categories.

Using a standardized research evaluation framework can serve two main functions: first, it allows comparison of evaluations, since they build from the same framework and categories of impacts; and second, it allows identification of unexpected outcomes of research, since a framework can help to ensure that all possible outcomes are investigated. Given that many funders of health research in Canada have already endorsed some type of logic model, the definition of a common framework to identify funding impacts in various categories that fit an acceptable logic model seemed reasonable.

2.1.a. Definitions

(Note: For a full glossary of definitions relating to this assessment, please refer to Appendix F, p. A311.)

For this assessment, the most important term for the panel to define is "return on investment." Traditionally, this term has been associated with the monetary return on research funding, but the Canadian Academy of Health Science has chosen a wider interpretation, which identifies health and societal benefits as a well as monetary ones (Canadian Academy of Health Sciences 2007). This broader definition prevents the use of evaluation findings to emphasize and focus strictly on economic impacts of health research. That decision introduces the potential of identifying more difficult to define but critically important 'value-based impacts' to society of health research investments.

While we are using such a broader definition of return on investment, it should be noted immediately that improvements arising in health, social changes, and economic benefits are likely due to the interaction of many factors and not just health research. That "attribution problem" of research on these health research impacts will be discussed more fully in Chapter 3.

The second most important definition for this assessment was that for the spectrum of potential health research "impacts". In this context, impacts were defined as the overall results of research on society and may include additional contributions to the health sector or to society (Canadian Institutes of Health Research 2005b). These include outputs and outcomes—two distinct concepts that are often confused. "Outputs" represent the tangible findings of research, and can be split into primary outputs (from the funded research, for example, publications and products), and secondary outputs (which arise because of the results of the research, for example, systematic reviews, guidelines, policies, etc.)

(Wooding, Hanney, et al. 2004). “Outcomes” are the effect that these outputs have on different stakeholders, either desired or unexpected (Development Assistance Committee Working Party on Aid Evaluation 2002).

”Health impacts” can be defined as changes in the healthy functioning of individuals (physical, psychological, and social aspects of their health), changes to health services, or changes to the broader determinants of health. “Social impacts” are changes that are broader than simply those to health noted above, and include changes to working systems, ethical understanding of health interventions, or population interactions. “Economic impacts” can be regarded as the benefits from commercialization, the net monetary value of improved health, and the benefits from performing health research.

2.1.b. Stakeholders Needs

As with any evaluation, understanding the mission of the organization and the desired impact of the evaluation is important in determining its success. In a sense, this is also true of establishing a useful health research impact framework for multiple stakeholders. The needs and expectations of the funders and other stakeholders in health research needed to be clearly understood before such a framework could be developed for them. For this assessment, interviews with sponsoring organizations and stakeholders defined needs in three distinct clusters: framework development, metric development, and evaluation issues (Figure 4).

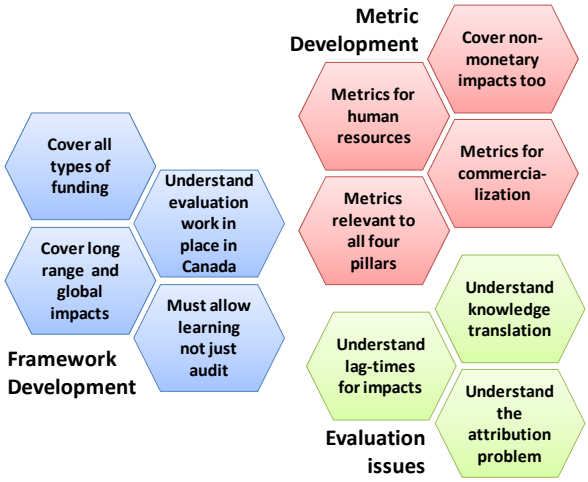


Figure 4. The three clusters of sponsor expectations for this assessment. These drove the assessment and the structure of this report: Frameworks (Chapter 2), Evaluation Issues (Chapter 3) and Metrics (Chapter 4).

There was a clear message from all sponsors that a framework must encompass all research funding (not just project grants), and that it should take advantage of the wealth of existing Canadian work on evaluating impacts of research (see Chapter 2 and Appendix C). Understanding the situation in Canada not only allows the best available information to feed into the assessment, but also increases the potential for synergy with current reporting systems (since a key aspect of any evaluation is to make cost-effective and integrated data collection decisions (HM Treasury, Cabinet Office, et al. 2001)). Any framework must be more than an audit tool: it must also support organizational learning, and be able

to cover the longer term and global impacts of research rather than just the immediate and local outputs of the research. As a tool covering all these outputs, the metrics (or indicators) must be wide ranging; covering all pillars and potential output categories.

Sponsors also indicated that this assessment should clarify the major problems in health research evaluations, including attribution, how to take lag times into account (covered in Chapter 3 and Appendix D), and whether effective evaluation can help determine how to best translate the knowledge produced by research.

Finally, they made it clear that they wanted an approach to either identifying or developing valid metrics that would be relevant to all four pillars of health research, to human resource issues in health research, to capturing impacts of commercialization as well as at least some important non-monetary impacts of health research.

2.2. A Review of Frameworks and Their Use

In addition to the rationale for frameworks noted above, an evaluation framework can also provide the opportunity to visualize the potential pathways to research impacts or the forms of impact that exist. As will be described below, that visualization can also help understand the system by which impacts are produced as thus improve it.

As reviewed already, there are many evaluation frameworks being used to help identify where and how impacts from health research occur. They all attempt to link downstream impacts to the research conducted, and all try to identify and attribute the impacts arising from research. They differ in approach (often focussing on either the pathway of research to impacts, or the categorization of impacts), and in their emphasis on different types of impact. They also differ in terms of whether they are designed for specific research project/program evaluations, or to allow routine data collection and evaluation of activities. They do, however, tend to lead to similar sets of generic indicators rather than to very specific indicators for individual evaluations.

Frameworks that focus on pathways to impacts are obviously designed to allow research to be tracked through to impacts, meaning valuable contextual factors are taken into account with regard to how research actually causes impacts. As such, these frameworks tend to be linked to evaluations for organizational learning, and tend to focus on some sort of logic model that links inputs to research processes, outputs, and outcomes. Examples of this type of framework include the research utilization ladder (Landry, Amara, et al. 2001) and the Weiss logic model approach (Weiss 2007).

Frameworks that classify impacts do not provide much information about how impacts came into being, but do allow a collection method for impacts that can be compared across evaluations. Hence, classification frameworks are well suited to evaluations that focus on accountability or advocacy, since they can identify a wide variety of impacts, including whether there are impacts linked to the organizational mission. Having categories of impact also creates obvious benchmarking potential with other evaluations (with the caveat, however, that different types of research have different sorts of impacts). Having impact categories also allows an organization to identify where it is *not* having impacts, as well as where it is. This is an important issue for accountability and advocacy, since it allows an organization to investigate more closely why it is “failing” to achieve under certain types of impact. Examples of this type of framework include the balanced scorecard (Kaplan and Norton 1992), the societal impact framework (van Ark 2007), the HTA organization assessment framework

(Lafortune, Farand, et al. 2008), the decision-making impact model (Lavis, Ross, et al. 2003), and the research impact framework (Kuruvilla, Mays, et al. 2006).

Perhaps the most commonly used health evaluation framework in the world is currently the "payback framework," which combines both of the framework types described above. It features a logic model that allows tracking of research from initial ideas through inputs, processes, dissemination, outputs, secondary outputs, and final outcomes. It also employs a multidimensional categorization of research impacts, which runs parallel to the logic model, and contains five categories, originally identified as: knowledge production; research targeting and capacity building; informing policies and product development; health and health sector benefits; and broader economic benefits (Buxton and Hanney 1996). Research is followed as a narrative using the logic model, while collecting impacts as they arise and assigning them to the appropriate category. As a result, this particular evaluation framework can be used for learning about how impacts arise, and for accountability or advocacy on the types of impacts that have arisen from research.

As already introduced briefly, several types of research framework are currently in use in Canada. Logic model frameworks are being used by a number of provincial funders and NAPHRO (Beaudet 2007). Others are using a logic model with additional framework aspects, such as the "action-reflection" approach to implementing evaluation findings used by the Saskatchewan Health Research Foundation (Saskatchewan Health Research Foundation 2007). Still other funders are using versions of the balanced scorecard (University Health Network 2008) as a process tool, not one for organizational learning. CIHR's version of a new payback framework, modified from the payback model, combines a logic model and "categorization of impacts" approach. It allows evaluations to be performed according to CIHR's organizational logic model, and provides multiple categories for collecting and ordering data on research impacts (Canadian Institutes of Health Research 2005b). A number of other organizations are collecting data that can also be considered when discussing evaluation in Canada, since these data inform indicators as well (Canadian Institute for Health Information 2008a).

To date, evaluation frameworks in Canada have generally been used for specific evaluations, rather than routine data collection. As of 2008CIHR's payback framework is beginning to evolve toward collecting routine data, while CIHI and Statistics Canada already collect large amounts of routine data short-term, for long term evaluation (impact) purposes. Outside Canada, there are also many frameworks being developed and used. For example, the payback framework in use at CIHR is also being used by organizations in Ireland and the U.K., and impact categorization forms a large part of evaluation frameworks in the Netherlands, Australia, and at the Wellcome Trust in the U.K.[7] For those who are interested, the different frameworks mentioned above (both conceptual and in use) are discussed in more detail in Appendix C, p. A232.

In terms of data sources, many international comparisons of research access information through the OECD and other international data repositories to make conclusions. At the national and organizational level, evaluations that attempt to make farther-reaching impact assessments use assumptions to derive information for what they hope will be more robust evaluations. Not least among these assumptions is the attribution of a variety of impacts to health research by implication as

[7] Theoretical, national, and international evaluation frameworks are covered in more detail in appendices A and C.

opposed to definitive proof. This issue is discussed in much more detail in Chapter 3. It will suffice to say that attribution is a major issue that has plagued many projects, and has been a major criticism of the "Exceptional Returns" studies undertaken in the USA and Australia (which 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 $3 million value of a life are regularly discussed), the assumed time lags between research and impacts (which are known to be long but variable), the ways in which impacts sum (the magnitude of various contributions to an impact), and the ethically sound need to avoid avoidance of double-counting of research impacts.

Contributing to some of these issues is the lack of funder balance in health research evaluations. Considering that most evaluations are conducted by research funders themselves with very little communication or collaboration, it is no surprise that there is not a balance of different research funders in evaluations. Attempts by countries to evaluate their R&D efforts, however, regularly focus on one type of research funder, to the detriment of the others (usually focusing on the wider impacts of public funding, and the sales or turnover of the private sector, while often ignoring the private not-for-profit sector entirely). The most recent work in the U.K. for the U.K. Evaluation Forum on the economic impact of cardiovascular 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).

Based on all of this background, improving current frameworks requires an analysis of their strengths and weaknesses, and the trade-offs that might prevent them from addressing all evaluation issues (a concept also discussed in Appendix A, International Evaluation Frameworks section, p. A147). By understanding what has worked well and what is commonly questioned, we can add value to any system and indicators for Canada in two major ways:

1. We can ensure that we use the best existing indicators and not replicate the shortcomings of other evaluation methods.
2. We can identify where a new framework and/or indicators can add most value to the science of "research on research."

2.3. Developing a Health Research Evaluation Framework for Canada

It is important to understand what current evaluation methods can and cannot do in order to be clear about what value a new approach might add. A review of the literature points to a rich database of potential evaluation tools and methods to potentially capture the "ROI" of health research, but it is necessary to classify those approaches in a way that makes them easier to understand and use. As per our charge from sponsors, that is our intent in this synthesis of frameworks: to provide one source for current models captured within a new framework that will help those trying to determine the ROI of health research, and to make the database behind its development as easy and "user-friendly" as possible.

Regardless of the approach taken to create such a "simple framework," we must make users recognize the many complicated interactions that produce and reinforce changes among multiple stakeholders once any new knowledge is created. Thus, any new framework must make it easy for the user of the framework and indicators to identify what might work for their unique purposes, while recognizing

and coping with the multiple levels of complexity inherent in a multi-funder, multi-provider, multi-stakeholder system.

2.3.a. Identifying What Should be Modelled

It is essential to ensure the correct factors are taken into account when modelling the Canadian health research system; but considering all aspects that could impact upon the outcomes of health research funding would be prohibitively complex, defeating the very aim of creating a model, which is to provide a simplified and utilizable version of events in the real world.

The first aspect to consider when assessing the outcomes of Canadian health research is the *expected* outcomes of the research. As reviewed in Chapter 1, this can ultimately be seen as an improvement in health and a corresponding improvement in wealth. For example, the CIHR mandate states that its role is to improve knowledge, but also to translate it into "improved health for Canadians, more effective health services and products, and a strengthened Canadian health care system."[8] Wealth is identified as important by a number of research funders, particularly those in the private sector, where a return on investment is vital in order to satisfy basic business rules of profitability. It is not just the private sector, however, that sees the value of creating wealth through health research. As in the CIHR mandate, more effective health services and a strengthened health care system can go hand in hand with improved economic outcomes for the nation and individuals.[9]

Since so many complex factors affect health outcomes, it is important to place any framework for research outcomes firmly within the existing frameworks for health determinants (for a summary of determinants and frameworks, refer to the WHO discussion paper, "Towards a Conceptual Framework for Analysis and Action on the Social Determinants of Health," Solar and Irwin 2006).

For the purpose of our work, we are interested in identifying where health research can impact upon health, with a health determinants model providing the important check on where research can reasonably have an impact on outcomes. We are not attempting to create a new model of the determinants of health. The work of Evans and Stoddart (1990), placing the determinants of health in a Canadian context, and that of Dahlgren and Whitehead for the WHO (Dahlgren and Whitehead 1991), has shaped how we have viewed the determinants of health for our framework. Figure 5 allows us to incorporate all the aspects of the Evans and Stoddart predictive model, while maintaining a simplified understanding of the factors that research can affect to affect health outside of the health care system (commonly viewed as the way in which health research goes on to impact upon individuals).

[8] Other funders of health research also have clear health goals, with charities and provincial level funders also expressing their desire to improve health. For example, the Saskatchewan Health Research Foundation has as its vision, "To build a healthy Saskatchewan through health research" (Saskatchewan Health Research Foundation 2008), while HSFC explicitly states that its role is to eliminate heart disease and stroke (Heart and Stroke Foundation Canada 2008).

[9] Although this is not always the case, as new health advances can end up costing the health system more. For example, new cancer drugs can often raise the cost of providing health care (Khoo, Colucci, et al. 2007).

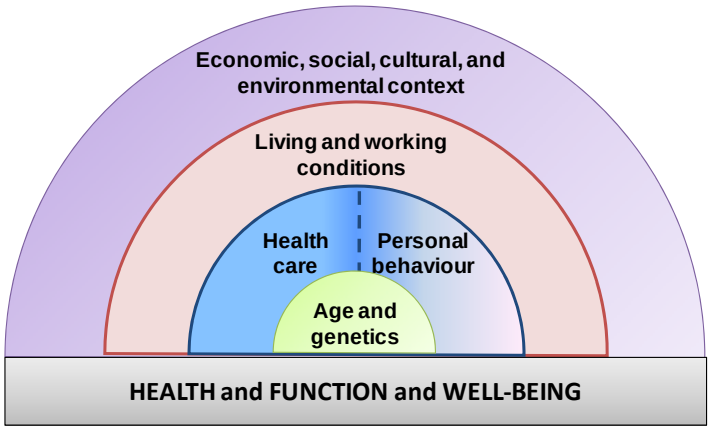


Figure 5. Systems map of determinants of health

This systems map does not show how research can impact upon health (or indeed any other aspect upon which research can have impacts). For that purpose, we require a framework that can follow the flow of research findings from all aspects of health research to all potential impacts.

2.3.b. Building a Framework for R&D Uptake

Developing an evaluation framework is a process that needs to take into account the intervention being evaluated. In our case, this means the entire Canadian health R&D effort (to include any element in isolation and all elements considered together). Since this has not been the focus of other, previous frameworks, it is unsurprising that no single framework currently explains adequately the impacts arising from all Canadian health research. In developing a framework for Canada as a whole, certain key factors from current frameworks need to be taken into account: the identification of target audiences for research findings, logical pathways to outputs and outcomes, a wide understanding of potential outcomes of health research, and categorization of impacts. While individual frameworks capture aspects of these – for example, the decision-making impact model and the research impact framework identifying target audiences but not providing a wide understanding of potential outcomes (Lavis, Ross, et al. 2003, Kuruvilla, Mays, et al. 2006) – each is focused differently to best evaluate specific aspects of research outcomes. In order to have a framework that encapsulates all of the different factors identified and is designed for the Canadian health R&D system, the panel has produced a CAHS framework that builds on the extensive evaluation experiences of the previous frameworks.

Perhaps the two most commonly used frameworks in R&D evaluations are logic models and the balanced scorecard. Logic models cover a wide range of (often separated) evaluation frameworks, since logic models themselves merely provide a way to visualize the inputs, processes, and outputs/outcomes of a program of R&D. The balanced scorecard differs from the logic model by providing a framework based on the main roles of the organization (originally the balanced scorecard looked to investigate outputs in four areas: financial, internal business process, customer, and learning and growth) (Kaplan and Norton 1992).

In terms of Canadian health research, there has been a clear trend toward the use of logic modelling as the basis for evaluation frameworks. CIHR's use of the payback framework, CHSRF's use of logic modelling, and the logic model approaches of the Alberta Heritage Foundation for Medical Research and Manitoba's Health Research Council (Canadian Institutes of Health Research 2005b, Canadian

Health Services Research Foundation 2005, Birdsell and Matthias 2001, Birdsell and Asselbergs 2006) all show that there is strong support for the logic model approach. The balanced scorecard has been used more predominantly in the USA and among private sector R&D providers (Osama 2006, Modell 2004, Bremser and Barsky 2004), but has been used in Ontario to evaluate health practice (Woodward, Manuel, et al. 2004).

At a conceptual level, there seems to be no reason why the logic model and balanced scorecard approaches should not be compatible, and it is our view that the two systems are merely evaluative tools that look at different sides of the same coin. For example, where a balanced scorecard approach looks to categorize the outputs and outcomes of R&D into the most important strategic directions for the organization, the logic model looks to categorize the same outputs and outcomes into "stages" of the intervention or the mission of the organization.

Since the prevailing strategy in Canada is for logic models, we support the payback framework approach, modified to capture the range of measures we believe to be important in the development of indicators. It is important to stress again that the framework itself is not the output of this work: the framework provides a logic within which to situate performance indicators that can best provide insight into the successes of, and challenges facing, the Canadian health research system. Also, using a logic model approach allows us to model the progress of health R&D toward the many types of outputs and outcomes that can occur, without prescribing for any individual funder of research how their specific type of health research program *should* progress.

The payback framework has been incorporated here because it has been adopted by CIHR and a number of other international research funders, and because it provides an opportunity for a multidimensional approach to research evaluation, incorporating both the flow of R&D toward outputs and outcomes, and the typology of impacts that can occur from R&D (see Appendix C, Describing the Payback Model, p. A232). The payback framework was originally designed for health services research (Buxton and Hanney 1996), but has since been adapted to address health technology assessments (Buxton and Hanney 2005), research in specific disease areas (Wooding, Hanney, et al. 2005; Hanney, Mugford, et al. 2005; Hanney, Home, et al. 2006), and has been adapted to all four "pillars" of health research by CIHR (Canadian Institutes of Health Research 2005b). This adaptability makes it an ideal framework on which to base a complex task, such as evaluating all the potential players in the health research market in Canada. Adaptability comes at a cost, however. In any evaluation using modified versions of the framework, it is important to maintain consistency in the ways that data are collected, particularly in identifying which impact category (payback category) particular outputs and outcomes fall into. Without this consistency, there is a danger that different versions of the framework will not be comparable.

There is also an element of pragmatism here, which we openly acknowledge. If the major funders of research in Canada are already moving towards an internationally used evaluation strategy, it would be both uneconomical and unwise to attempt to completely re-draw the framework. The cost-effectiveness of implementation and likelihood of use were also important considerations for us. Further, the payback framework provides a level of adaptability in its ability to address the research impacts of many different types of research—something that is particularly pertinent to our cause in trying to provide a framework that works with all four pillars of health research in Canada, and with the different funding mechanisms used by the myriad research funding organizations in the country.

Figure 6 illustrates the logic model that has been created by our panel. It clearly focuses on the ways in which health R&D can have impacts. This framework differs from the payback framework (shown in Appendix C, p. A226) in that the payback framework itself contains opportunities to investigate topic identification, selection, inputs to research, and the research process itself. These are less important to articulate in a framework that is designed to identify the most appropriate indicators for impacts of health research (products of the research process, as opposed to inputs) across Canada.

As shown by the top arrow, the framework moves from the underlying reasons for health R&D funding through the initiation and diffusion of health research to the various different levels of outputs and outcomes. It looks explicitly at the *processes* by which health R&D can go on to have impacts (that is, outputs and outcomes). This framework mimics that of the payback model (shown at the bottom of Figure 7 in red text), but focuses more on the secondary outputs and outcomes of research, since these areas are the most difficult in which to identify performance indicators.

The initiating stage of the framework identifies the health needs and the costs of research (driving factors in deciding health research topics). The second stage contains the actual R&D that is performed in Canada (as well as a box showing the parallel global research effort). This covers many aspects of research from identifying research topics and getting them funded (*selection* in the payback framework), through to the research process itself (*process* in the payback framework). In the Canadian Health Research box, we list the four pillars of health research as defined by CIHR (Canadian Institutes of Health Research 2007d): biomedical research, clinical research, health services research, and population and public health research. Having the four pillars explicit in the framework allows us to follow the most likely paths of research impact in any pillar. It also allows us to "personalise" the framework for each pillar by helping to identify the most appropriate indicators available for research in each pillar. We also identify "cross-pillar" research as an additional type occurring in Canada. Below the research activity box, we identify the research capacity of Canada, including human capacity, infrastructure capacity, and capacity built in education and research tools.

Research produces results and knowledge (*primary outputs* in the payback framework). Clearly, research results have no effect until they are made available; therefore, results must enter the knowledge pool in order to have impacts (*dissemination* in the payback framework). Equally clear is that there are a number of ways in which research can be disseminated (and, indeed, a number of ways to define "knowledge pool"). In this project, we refer to dissemination in its widest possible sense, in that it encapsulates any method of transferring information from the researcher to a wider audience (be they scientists, health practitioners, industry, government, or the public, for example). This also means that "knowledge pool" here has a very broad meaning, whereby any advance in knowledge by any group is considered a broadening of the knowledge pool.

The next stage of the framework follows the outputs of R&D to the groups where they are disseminated. Clearly, broadening the knowledge pool is an important aspect of R&D. For it to have an impact, however, disseminated R&D must cause a change in behaviour or decision making. This can occur through a two-way process involving collaboration and consultation between research users and researchers. In our framework, we have identified five groups upon whom R&D outputs can impact: the health industry, other industries, government decision makers at all levels, policy makers in research policy, and the public and specialist groups (such as advocacy groups). These groups correspond to *secondary outputs* in the payback framework: the level at which organizations or individuals take the outputs of R&D and use them in their own outputs.

The Framework for Health Research Uptake

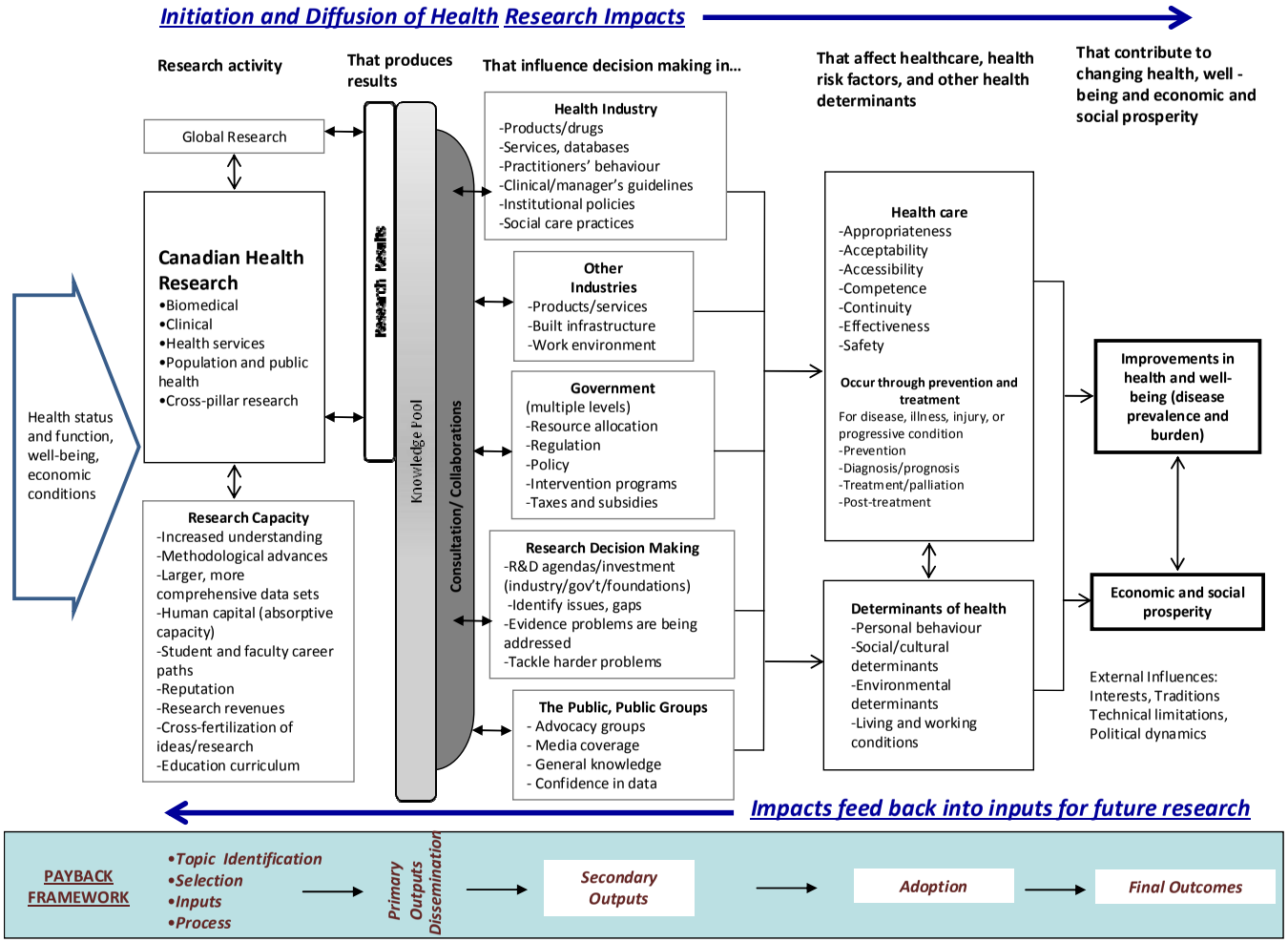


Figure 6. CAHS framework logic model of health research progression to impacts

As in the payback framework, in order for secondary outputs to lead to outcomes (for example, better health and greater wealth), they must be adopted by either an organization or individuals. This is tantamount to any changes in secondary outputs actually changing behaviour. This occurs in two distinct places: first in the health care system (including public health and social care systems), by improving quality, access and cost of treatment (for example, through changes to prevention, diagnosis, treatment, and post-treatment), and second, in the impact upon individuals, where adopting secondary outputs affects determinants of health (be those individual behaviours, environmental, or social changes).

Once R&D findings have been adopted and have changed behaviour, they then go on to have their *final outcomes* (in payback framework terminology) in changing the health and well-being of the population, and the economic and social prosperity of the country (and individuals/organizations).

It is important to remember that research findings are not the only reason that there are (or are *not*) changes to health, well-being, and other impacts identified in the framework; and we capture some of those external influences by identifying interests, traditions, technical limitations, and political limitations as reasons why outcomes may or may not come about. We also note on the framework that the nature of research findings having impacts is not linear. The lower blue arrow shows how impacts that arise feed back into the inputs for future research. For example, research that has a big impact on our understanding of cancer would have large consequences on new research being proposed.

Figure 6 shows the full logic model, with details of the factors that make up the different players and processes for each box. Throughout the rest of this document, however, for the sake of space, we use a simplified model showing only box titles that represent the top-level factors the framework takes into account.

2.3.c. Impact Categories

As mentioned with Chapter 2 above, the payback model incorporates both a multidimensional categorization of research impacts and a logic model of pathways to impact. Impact categories serve the purpose in the model of "bins" of research impacts, and in an evaluation framework that is used as an ongoing evaluation tool. These "bins" are the receptacles for indicators of research impact (identifying what should be collected in order to identify whether the expected impacts are being achieved).

The original payback framework contained five payback categories, which have been adapted in many payback evaluations conducted since (for example, Wooding, Nason, et al. 2007). Particularly relevant to our work, however, is the modification of the five payback categories by CIHR (Canadian Institutes of Health Research 2005b, Canadian Institutes of Health Research 2008b)(Table 1).

Table 1. Mapping of payback categories from the original categories to the most recent CIHR categories

Original Payback Category (Buxton and Hanney 1996)	**CIHR 2005 Categorization** (Canadian Institutes of Health Research 2005b)	**CIHR 2008 Categorization** (Canadian Institutes of Health Research 2008b)
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

To ensure the usability of our framework, we need to demonstrate that the impact categories also fulfil the required criteria for our model, which are:

- being useful to a range of funders/research types
- being compatible with what is already in place in Canada
- being transferable to international comparisons
- being able to identify the full spectrum of potential impacts

In order for the model to be applicable to a variety of research funders, what is captured by each category and how categories fit with our framework and with the four pillars of research must be clearly understood. To that end, we have modified the categories slightly from those used by CIHR to ensure that all potential impacts of Canadian health research can be captured: advancing knowledge, capacity building, informing decision making, health impacts, and broad economic and social impacts. The specific indicators that are most appropriate for these categories are discussed in Chapter 4, but below we outline the types of data that each category encompasses. For full details on the different indicators that exist in the world for each of these impact categories, refer to Appendix E, p. A284.

Our versions of the category definitions are built upon the use of different indicators in health research evaluation globally. Since an ideal evaluation should be a comprehensive assessment of impacts from an intervention (in this case, health research funding), we have identified subcategories in each impact category that should be investigated to ensure that the evaluation addresses all potential impacts (Table 2).

Table 2. CAHS framework impact categories and subcategories

Category	Level 1 subcategory	Level 2 subcategory
Advancing Knowledge	Research quality	N/A
	Research activity	
	Outreach	
	Structural	
Capacity Building	Personnel	N/A
	Activity funding	
	Infrastructure	
Informing Decision Making	Health related	Health care
		Public health
		Social care
		Other
		Health-related education
	Research related	Research funding
		Research policy
		Research education
	Health products industry	N/A
	General public	Advocacy groups
		Public education
Health Impacts	Health status	Morbidity
		Mortality
		Quality-adjusted mortality
	Determinants of health	Modifiable risk factors
		Social determinants
		Environmental determinants
	Health care system	Acceptability
		Accessibility
		Appropriateness
		Competence
		Continuity
		Effectiveness
		Efficiency
		Safety
Broad Economic and Social Impacts	Research activity	N/A
	Commercialization	
	Health benefit	
	Well-being	
	Social benefits	

Advancing Knowledge

This CIHR category includes new discoveries and breakthroughs from health research, and contributions to the scientific literature (Canadian Institutes of Health Research 2008b). In our classification of advancing knowledge, there are four subcategories of indicators:

- research quality measures, which indicate the "quality" of that research output (as defined by some measure of research quality)
- research activity measures, which measure the volume of outputs of research

- outreach to other researchers, which show how researchers interact in terms of research (both conducting and using research)
- structural measures of the research portfolio, which capture the way in which a research organization (be it a performer or funder) balances its portfolio of different research fields

In general, advancing knowledge indicators have centered on bibliometric analyses and surveying of researchers.[10]

Capacity Building

This category includes the development and enhancement of research skills in individuals and teams (Canadian Institutes of Health Research 2008b). It is interesting to note that CIHR explored the idea of rolling capacity building into advancing knowledge (Canadian Institutes of Health Research 2007b), showing the close relationship between advancing knowledge and building research capacity. This is mirrored in our logic model (Figure 6), which shows research capacity at the same level as research activity but, like CIHR in 2008, we felt it should be separated. We noted that capacity building has three subcategories:

- personnel, which covers the current capacity for research through research-related staff, the next generation of health researchers, and the ability of personnel to take on research findings (receptor and absorptive capacity)
- additional activity funding, which covers money brought in by current research activity (such as matched funding)
- infrastructure, which covers any aspects of research infrastructure required to perform research (including how infrastructure is linked to activity)

Informing Decision Making

In the CIHR system, this category includes the impacts of research in the areas of science, public, clinical, and managerial decision-making practice and policy (Canadian Institutes of Health Research 2008b). Informing decisions is the crucial step in how research goes on to wider outcomes. Decisions can be either evidence based (acting upon specific evidence), or evidence informed (acting upon a knowledge of the evidence, but not on a specific piece of evidence), and it is important to try to capture both of these concepts. Since decisions are made on the basis of a large number of influences, and since it is often difficult for even the person making the decision to identify what the major factors were in the decision, the way research informs decisions is an inherently difficult process to measure. As a result, there are often proxies for decision making in the indicators previously identified in this impact category.

[10] An important aspect to note here is the difficulty in measuring the *progress* of an advance in knowledge. For example, research on cancer may be five, 10, or 20 years from fruition, and most funding organizations would wish to know which. Outside of clinical trials performed on specific treatments/vaccines, etc., putting an exact number on the progress toward a desired health outcome is very difficult. This issue is also noted by Hage in Appendix A.

In an attempt to subdivide this category and allow the collection of indicators that cover all types of decision making, our classification has four subcategories, each with its own subcategories. They are outlined below:

- *health-related decision making*, which covers decision making in (a) health care, (b) public health, (c) social care, (d) other health-related decision making (for example, health and safety at work), and (e) health-related education (such as training of new health professionals and continuing health professional education for health professionals)
- *research-related decision making*, which covers (a) decisions about research funding allocations, (b) research policies, and (c) researcher education
- *health products industry decision making*, which has no associated subcategories
- *general public decision making*, which includes (a) the decisions of advocacy groups (such as patient groups), and (b) the way the public is educated about research

Health Impacts

In the CIHR classification, this category encompasses advances in prevention, diagnosis, treatment, and palliation when related to research (Canadian Institutes of Health Research 2008b). The difficulty arises, however, in how to relate findings to research. For example, CIHR identifies changes in health as an indicator, but then relies on "special studies" to link those to research findings (Canadian Institutes of Health Research 2005b). We have examined the best ways to identify health and health systems improvements, and have tried to link to research wherever possible.

Our categorization, based on the ongoing work of the Canadian Institute for Health Information (CIHI), splits health impacts into three major groupings: health status, determinants of health, and health system performance (Statistics Canada and Canadian Institute for Health Information 2008). Since research has effects by informing prevention, diagnosis, treatment, and post-treatment, these aspects are captured in the changes to health status and the performance of the system.

Our main health impacts groupings, with their related subcategories are outlined below:

- *health status*, which covers (a) mortality, (b) morbidity and (c) quality-adjusted mortality (a measure of health status that takes into account quality of life)
- *determinants of health*, which have been widely discussed and well covered. We have identified those that health research can have impacts upon: (a) modifiable behavioural determinants (such as smoking), (b) social determinants (such as health education levels), and (c) environmental determinants (such as air pollution levels)
- *health care system performance*, which covers the eight characteristics identified by CIHI: (a) acceptability (to service users), (b) accessibility (for different service users), (c) appropriateness (the service uses the best evidence to guide it), (d) competence (the service doesn't make mistakes), (e) continuity (service users get continuity of care), (f) effectiveness (outcomes from the service are the best possible outcomes), (g) efficiency (the best possible outcomes at an appropriate cost), and (h) safety (service users experience safe service) (Canadian Institute for Health Information 2008a). While these are not measures of health per se, they can be

impacted by health research and have secondary effects on health in many dimensions. We therefore felt that they are reasonable indicators of health research impact.

Broad Economic and Social Impacts

CIHR defines this category solely as "economic benefits," and divides it into three subcategories: commercialization of discoveries, direct cost savings, and human capital gains (Canadian Institutes of Health Research 2008b).

In our categorization, we also split the economic impacts into three subcategories, although somewhat modified from the CIHR choices. We also widened the category to include social impacts, since the framework needs to take into account all possible impacts from Canadian health research. Our broad economic and social impacts categories are described below:

- *research activity*, which is the benefit accrued through the action of research, as opposed to the outputs of research
- *commercialization*, which is the economic benefit accrued through the sales and revenues of commercialized research findings (such as sales of pharmaceuticals and the value of spin-off companies)
- *health benefit*, which is a measure of the net benefit of improving health; it includes direct-cost savings, and uses a value of health improvement (not human capital) and compares it to the cost of implementing that health gain (costs to the health system and of producing the research)
- *well-being*, which covers the many aspects that make up well-being, including levels of happiness and feelings of isolation (Joseph, Linley, et al. 2004, McWhirter 1990);
- *social benefits arising from health research*. For example, there are social benefits beyond improved exercise regimes for individuals from health research that suggests more playing fields or organized sports (both of which are conceivable recommendations from research to reduce obesity, diabetes, and heart disease). These social impacts include an improved environment and improved social capital through team membership. It is important to remember that these outcomes do exist and are worth trying to capture separately from the economic measures.

2.3.d. Impact Categories and the Logic Model

As mentioned, the relationship between impact categories and sections of the framework is not simple: it is possible for an impact to occur at any point in the framework (Nason, Janta, et al. 2008). This does not mean, however, that impact categories do not commonly link to sections of the logic model. For example, broad social impacts are most likely to occur at the final outcome stage of the model, with knowledge impacts most likely to occur at the primary output and dissemination phases. The links between the logic model and the impact categories are covered in more detail in the next chapter, where we explain how to use the framework to evaluate health research and the issues that evaluations are likely to encounter.

3. Chapter 3: Strategies for Using the Framework

Having developed a framework and impact categories, it is important to have guidelines for using them appropriately, since evaluating health research needs to be as ethical and accountable as the research itself. This requires understanding both how to use the framework, and how to avoid misusing it.

3.1. Using the Framework Appropriately

Because the framework consists of two elements (a logic model and impact categories), it is necessary to understand how the two interact. Impacts can potentially occur anywhere in the logic model, but more often than not, they occur at the same levels, helping to identify where indicators should be found, and how impacts link to the research itself. **Figure 7** shows the main areas of overlap between logic model sections and impact categories. Within the matrix, we show hypothetical examples of impacts from research into the molecular mechanisms of cartilage degeneration.

FRAMEWORK / **IMPACT CATEGORIES**	Health R&D	Primary Outputs/ Dissemination	Secondary Outputs	Adoption	Final outcomes
Advancing Knowledge	• New molecular technique developed	• Publication of research results in a journal			
Research Capacity	• Research PhD gained by team member		• Further research in Industry		
Informing Decision Making		• Discussions between researchers and pharma define direction of pharma research	• Pharma company initiates research program to develop a drug	• Drug developed passed for use by the health system	
Health Impacts				• Adoption by the health system causes increased cost of drugs • Decreased re-admission for condition	• Reduced condition burden in the population
Broad Economic and Social Impacts			• Employment in the pharma company	• Sales of drugs by pharma	• Improved workplace productivity • Social and economic benefit of "wellness"

Figure 7. Illustrating the overlap of the impact categories and the framework. Yellow shading represents predominant areas of overlap between sections of the framework and impact categories.

The framework is the vehicle for collecting data in a co-ordinated and replicable fashion, and provides an insight into where to expect impacts from health research to occur. Organizations using the framework will tend to focus on certain impacts that link to their mission. This in turn suggests which indicators should be used by different organizations for data collection. The categories provide a way to partition collected data using the framework and various evaluation methods. It is worth stressing

that organizations should tailor evaluations to fit their organizations, but should not ignore their undesirable impacts by addressing only positive impacts in their frameworks.

The framework can also be used for prospective evaluations to help identify where funding should be apportioned, since it can help to identify where impacts are likely to occur. As noted by Hage in his paper on evaluating health research outcomes at the meso-level (refer to Appendix A, p. A79), "recognizing the differences between treatment sectors allows policy makers to more correctly discern which areas of investment in medical research are more likely to have the highest pay-off measured in terms of various social and economic benefits provided that they can estimate the amount of time and effort needed to achieve a particular research finding."

3.1.a. Avoiding Misuse of the Framework

Misuse of the framework presents a number of challenges, not least to the validity of the evaluation being undertaken. Appendix A contains a paper on ethics in research evaluation by MacDonald and Knoppers (p. A121), an issue that is an important consideration in how the framework, including impact categories, is used.

Not only must evaluations take into account "desirable ethical impacts" (discussed in more detail by MacDonald and Knoppers), but the evaluations themselves must be accountable and open in their use of particular indicator sets and evaluation methods. When evaluations meet these criteria, it is possible to avoid some of the major pitfalls in impact evaluation for research, such as double counting of impacts (for example, the economic benefits from improved health due to cardiovascular disease could also be considered attributable to diabetes research), and the Halo Effect (the collection of only impacts that show the research in a positive light).

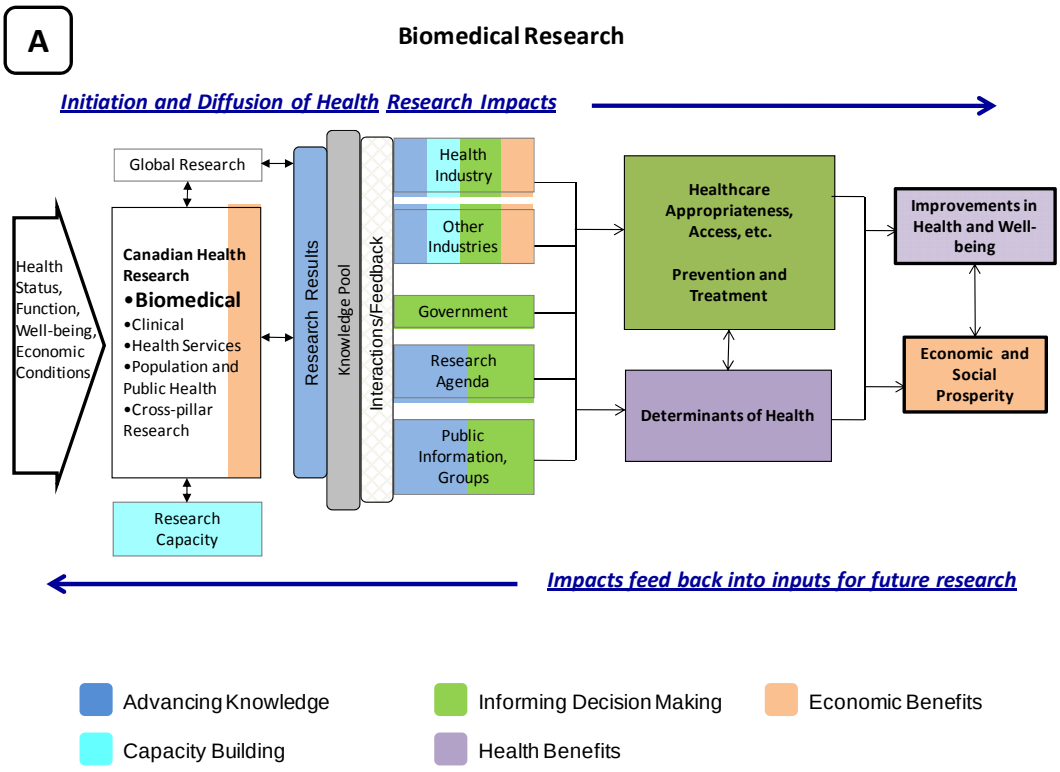
Double counting of research impacts is a very common issue in health research evaluation, and is difficult to account for, since health impacts and the monetization of health impacts are inherently double counting. Having the framework in place to identify the flow of impacts does, however, allow us to identify where double counting occurs. Alongside the framework, the impact categories also play a part in negating double counting. By not summing impacts across categories, it is possible to collect the information that best represents each category without worrying about double counting. For example, collecting licensed patent data as both a measure of informing the decisions of industry and also as a monetary measure of commercialization (economic impact) would not represent double counting, as long as impact categories are not summed.

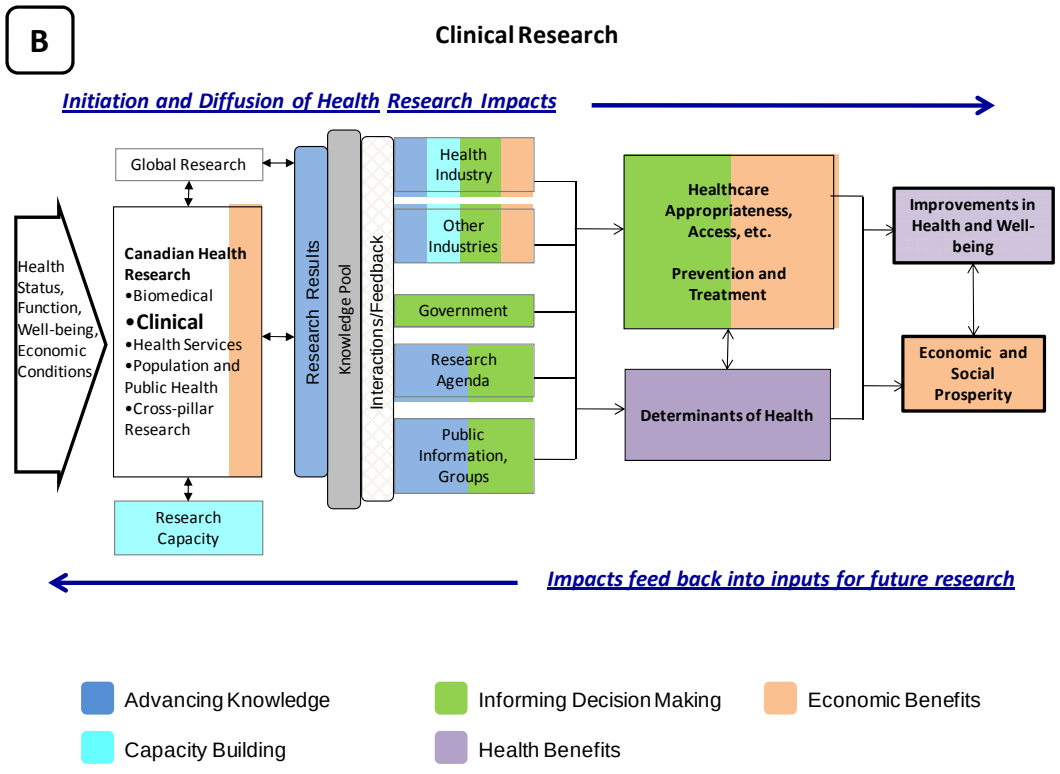
The Halo Effect stems from difficulties in identifying all research impacts (not just positive ones). Having a logic model for the evaluation means that all avenues of impact can be investigated, while having impact categories means data on specific indicators can be collected. These aspects of the evaluation framework should allow evaluators to investigate all impacts arising from the research in question.

In using the framework ethically, there are also considerations regarding informed consent for data collection and privacy of data (Appendix A, p. A121 - MacDonald and Knoppers). Collecting data from a variety of different funders and institutions and creating large data sets for use in analyzing Canada's health research impacts makes this an issue for incorporating national datasets. These issues need to be addressed when designing data collection methods.

3.1.b. The Four Pillars and the Framework

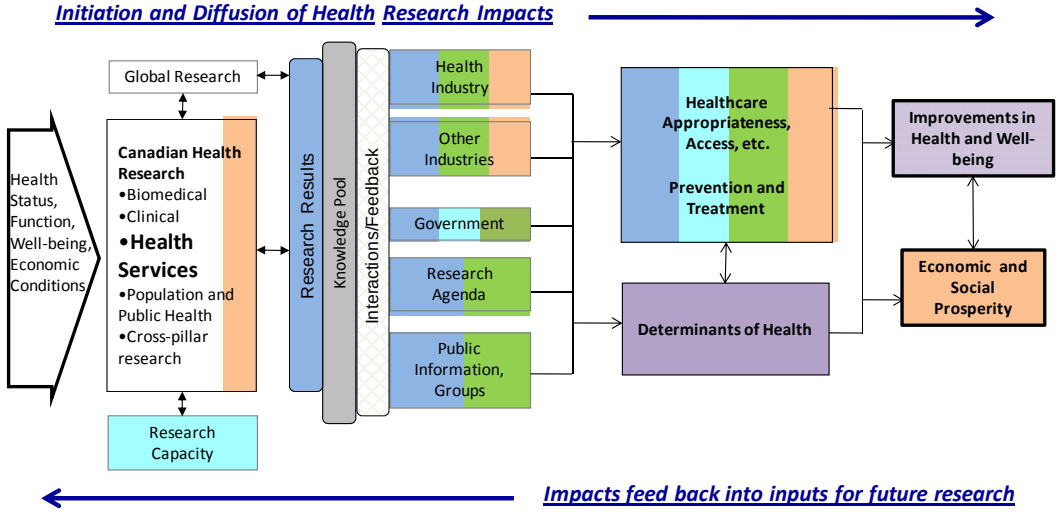
Any general framework that can be applied to all Canadian health research must be adaptable to the different impacts and pathways to impacts in each of the four pillars of health research, and in cross-pillar research. The logic model explicitly builds in the four pillars to ensure that we understand how each one progresses to impacts. Then, by cross-mapping impact categories for each pillar, we can create a specialized framework to show where the majority of impacts are likely to fall within that pillar (Figure 8). The areas of impact in this figure are not exclusive of impacts in other domains, but rather are intended to show where most impacts are likely to be found and thus where indicators or metrics for those impacts should be concentrated.

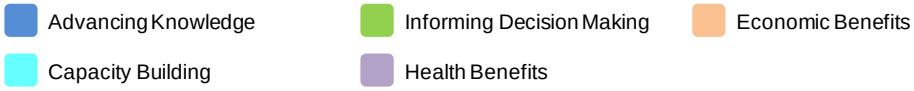

A
Biomedical Research
Initiation and Diffusion of Health
Research Impacts
Global Research
Health Status, Function, Well-being, Economic Conditions
Canadian Health Research
•Biomedical
•Clinical
•Health Services
•Population and Public Health
•Cross-pillar Research
Research Capacity
Research Results
Knowledge Pool
Interactions/Feedback
Health Industry
Other Industries
Government
Research Agenda
Public Information, Groups
Healthcare Appropriateness, Access, etc.
Prevention and Treatment
Determinants of Health
Improvements in Health and Well-being
Economic and Social Prosperity
Impacts feed back into inputs for future research
Advancing Knowledge
Informing Decision Making
Economic Benefits
Capacity Building
Health Benefits


B
Clinical Research
Initiation and Diffusion of Health
Research Impacts
Global Research
Health Status, Function, Well-being, Economic Conditions
Canadian Health Research
•Biomedical
•Clinical
•Health Services
•Population and Public Health
•Cross-pillar Research
Research Capacity
Research Results
Knowledge Pool
Interactions/Feedback
Health Industry
Other Industries
Government
Research Agenda
Public Information, Groups
Healthcare Appropriateness, Access, etc.
Prevention and Treatment
Determinants of Health
Improvements in Health and Well-being
Economic and Social Prosperity
Impacts feed back into inputs for future research
Advancing Knowledge
Informing Decision Making
Economic Benefits
Capacity Building
Health Benefits

C

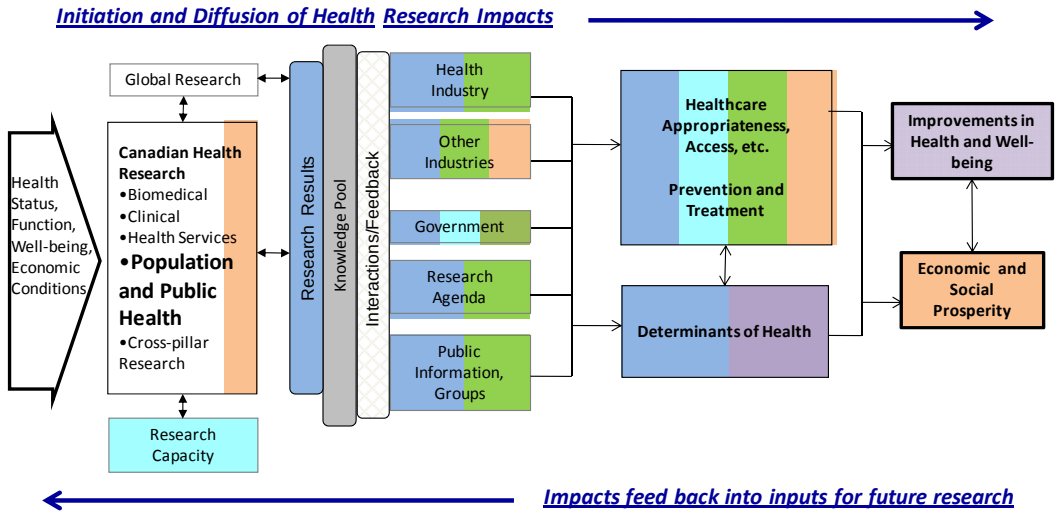
Health Services Research

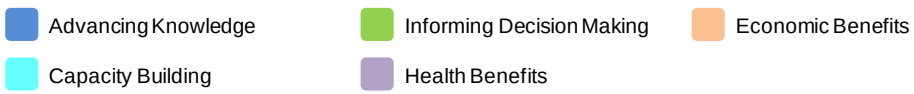




D

Population and Public Health Research





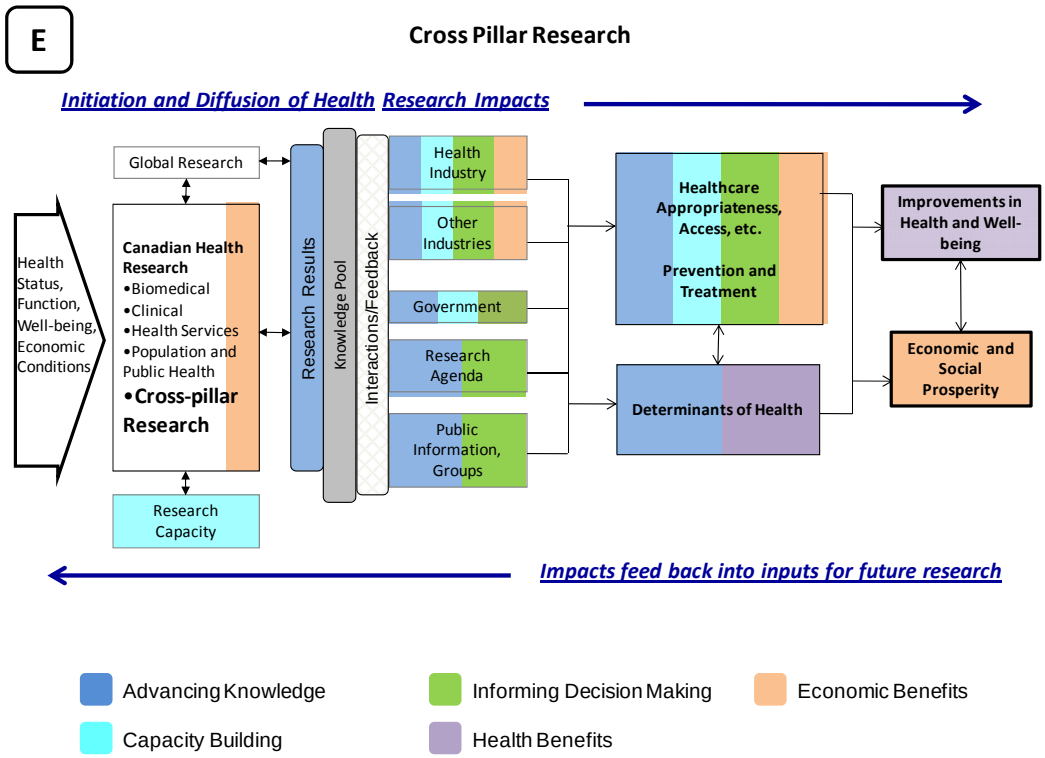


Figure 8. Impacts for each of the four pillars of research (A = Pillar I: basic biomedical research; B = Pillar II: applied clinical research; C = pillar III: health services research; D = Pillar IV: population and public health research; E = research across pillars)

In Figure 8, for biomedical research (box A), there are advancing knowledge impacts in the research findings for industry, for those forming the research agenda, and in the public. Capacity building occurs in the research community (academic and industry sectors). Informing decision making occurs for industry, policy makers, the research community, the public, and the health care system. Health impacts occur for determinants of health and on the health of the population. Economic and social impacts occur because of the activity of research, the commercialization of research by industry, and through improvements in the health of the population and changes to determinants of health such as the environment.

For the other pillars of research—clinical (box B), health services (box C), and population and public health (box D)—there are similar patterns of impacts, but with specific differences in each pillar. For clinical research (box B), there are additional economic impacts likely in the health system through changes to clinical practice and potential new drugs and devices. For health service research (box C), capacity can be built for policy makers (receptor capacity), but this is less likely for industry (absorptive capacity). Also, health service research directly affects the health care system (in its broadest sense), and can advance knowledge, build capacity, inform decisions, and make changes to the functioning of the health system. Population and public health (box D) has impacts similar to health service research, although the health care system changes relate to the public health system. Population and public health can also advance knowledge directly related to determinants of health.

Cross-pillar research (box E) can have impacts in any of these areas, since it can include research from any pillar. For example, the Canadian longitudinal study on ageing links research from all four pillars, from basic research on cellular ageing through to population studies on the interaction of ageing with lifestyle and environment (Canadian Longitudinal Study on Ageing 2007). To evaluate such an initiative requires an understanding of the impacts from each pillar, as well as how pillars interact to produce results, based on the way research informs other pillars of research.

3.2. Costs of Evaluation

It is not just research that must be accountable: any evaluation must also show value for money. This is inherently difficult to address since, as we have discussed, the type and extent of evaluation undertaken by an organization is entirely dependent on the sorts of questions that the evaluation intends to answer. However, previous work has suggested that spending between 1 and 5% of the research budget to evaluate outcomes is not unreasonable, and this can be used as a "rule of thumb" (Maredia, Byerlee, et al. 2000; Gibbons and Georghiou 1987). As a comparator, Human Resources and Social Development Canada spent between $150,000 and $500,000 on each program evaluation it undertook in the first half of this decade (Laurier 2005). It is important to note that it costs money to improve data collection, and that there are trade-offs between the quantities and quality of information collected and the cost of accessing that information. By funding research evaluation, however, it is not only possible to identify impacts from specific funding, but also to better understand the causal pathways that lead to impacts—improving the framework for future evaluations.

One way to reduce the cost of research evaluation is to ensure that the evaluation techniques are validated methods that can identify answers to specific questions. There are already a number of systems in place in Canada that can be built upon to ensure that funding for evaluation is used wisely and not spent on "re-inventing the wheel." Another way to reduce the cost of evaluation is to use data that is already routinely collected (HM Treasury, Cabinet Office, et al. 2001). This could represent national data collected through Statistics Canada or routine organizational management data (see "Data Collection," page 3). As an overview, the World Bank has produced documentation about how to perform impact evaluations under budgetary constraints that provides more details on reducing costs (Bamberger 2006).

3.3. Issues in Evaluation: Attribution, the Counterfactual, Time-lags and Levels of Aggregation

Evaluating the impacts arising from research is far from straightforward, even with a framework and designated impact categories. Four main challenges present to any evaluation: attribution, the counterfactual, time-lags to impacts and aggregation levels for evaluation.

Attribution, the ascription of a causal link between observed (or expected to be observed) changes and a specific intervention (Organisation for Economic Co-operation and Development 2002), has long been a problem for impact evaluation. Attribution can be considered at a number of levels. For health research evaluation, we can identify four main levels that become progressively more problematic (Smutylo 2001):

- attributing impacts to *research findings*, showing that impacts are the result (wholly or partially) of *any* research findings

- attributing impacts to *Canadian* research findings specifically (as opposed to those from other countries)
- attributing impacts to Canadian *health* research findings (as opposed to those from other fields of research)
- attributing impacts to *specific* research findings (such as those from a funded study)

Attribution is made more complicated by the number of exogenous factors that could also be involved in any impact. As a metaphor, the impact of throwing a single pebble into a pool near a fishing float is clear to see: ripples cause the float to move. However, as more and more pebbles are thrown in, many ripples interact, causing complex impacts on the float that are not easily attributable to any single pebble. With some impacts, attribution to specific research can be very clear, such as the 50% reduction in Sudden Infant Death Syndrome (SIDS) deaths in the U.S. as a result of the "Back to Sleep" campaign, which was built on research into infant sleeping behaviour (National Institute of Child Health and Human Development 2008). In other cases, attributing to research can be very difficult, for example, the improvements in respiratory health due to smoking bans.

In conducting evaluations, there is a danger of over-attributing findings to research—something that risks double-counting of impacts, since multiple research factors will all be considered to have been responsible for the impact. While this is the most common risk with attribution in research evaluation, there is, in some situations, also a real risk of under-attribution. For example, while 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.

Addressing attribution is complex and should ideally be done using a variety of methods (triangulation). The framework itself provides one such method, since it allows an identification of the narrative of the impact. By tracing the impact back through the framework, it is possible to identify which research has been involved. This approach can be expensive, as it requires case studies to understand the way research has been involved in the impact. It also runs the risk of providing unrepresentative data if the case studies are not chosen carefully. It is important, therefore, when creating a raft of case studies to do so using a selection framework that ensures a representative sample, dealing with several external threats to the validity of the evaluation (Program Evaluation Branch, Office of the Comptroller General 1991). Having a large collection of representative case studies provides even greater analytical power, and having case studies performed according to the same framework would allow such a bank of studies to be built.

An alternative approach to addressing attribution is to use multivariate regression analyses, which try to partition the impact between its various causes based on various explanatory variables. This approach was used by the OECD in "The Sources of Economic Growth in OECD Countries," a multivariate growth (GDP) regression for 21 OECD countries from 1971 to 1998, which identified the major causes of GDP growth and the level of impact of each (Organisation for Economic Co-operation and Development 2003). Regression analyses, however, rely on the quality of data and the explanatory variables included; with poor data or missing variables they can provide misleading attribution. By employing a number of techniques to address attribution, it may be possible to link impacts to research more closely and with better certainty.

Where attribution is not possible, some evaluations have chosen to address contribution to impacts. This has been used in some international development evaluations, where it was considered "good enough" to show that a development intervention had a role in the impacts identified (Earl and Carden 2002). For health research, contribution can provide useful insights into the way research has impacts, but is less useful for identifying what can further improve contribution to future impacts.

Understanding the impacts of research funding relies not only on attributing impacts to research, but also on understanding what would have happened if the funding had not been put in place. Identifying this counterfactual is a problem for all evaluations, and is most commonly addressed using a comparison (or control) group for the evaluation (Organisation for Economic Co-operation and Development n.d.). The gold standard for evaluation with a comparison group is the randomized control trial (Duflo and Kremer 2003), but this is not often 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, it is important to identify baseline measures and context in order to understand 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.

Time lags to research impact present a sizeable issue for any evaluation. Different research has different time lags to impact. For example, we know that health services research is more likely to have a fast effect on the cost effectiveness of health care, while basic science may take a very long time to impact on health. As well, impacts can persist for different periods and change over time.[11] An example of this is the recent studies into hormone Replacement therapy (HRT) that have shown new dangers associated with long-term treatment that were not previously known. This situation clearly shows the importance of understanding time lags in addressing impacts.

In the U.K., health research evaluations have attempted to address time lags through the use of clinical guideline studies that identify the time from publication of research to its inclusion in clinical guidelines as the minimum time to health impact (Buxton, Hanney, et al. 2008). This is clearly problematic, since it only applies to impacts that occur through clinical guidelines, but it does provide an approach for estimating times to impacts. In agricultural research evaluations, some groups have used infinite lag modelling, an econometric technique that requires large amounts of data to model the potential different time lags that exist for different impacts (Alston and Pardey 2001). In reality, however, the data for such a technique is not currently available. The use of finite lag modelling (applying a maximum time lag for impacts) is a more realistic approach for health research, and is an approach that fits with the one used in the recent U.K. study (Buxton, Hanney, et al. 2008).

Aggregation levels vary from evaluating individual researchers through to evaluating Canada as a whole.[12] These different levels of evaluation bring different challenges, and it is often the case that a

[11] Having a better understanding of the time lags to research impacts can help with an understanding of the progress towards the ultimate goal of the research (see footnote **10**)

[12] It is important to note that aggregation levels identify the level of the group being evaluated, not the level of the group performing the evaluation.

particular indicator is useful at some levels of aggregation and not others. This assessment has identified five levels of aggregation, based on the needs for evaluating each level (Box 1). Metrics for each level of aggregation are not always different, since summative indicators can often provide important information and aggregating indicators can help reduce the data collection burden. The levels of aggregation identified below mirror those developed by Hage in his paper on evaluating health research at the macro-, meso-, and micro-levels (see Appendix A, p. A79).

Box 1. Levels of aggregation for evaluation

1. **Individual**: Assessing the impact of the individual can be very difficult the further impacts occur from the research findings (the attribution issue).
2. **Research group/grant**: The impact of the group or grant (or any funding block) is slightly more straightforward than the individual, since the number of impacts is likely to be larger and more easily identifiable.
3. **Institution/department**: At the institution or department level, there are a large number of impacts available to evaluate, and the link to specific research findings is less important, since we need only locate research at a coarse level.
4. **Funding agency**: For funding agencies (or any funder, including industry), there is a need to show impacts from their own funded research. Attributing to this level may be difficult for health or economic benefits.
5. **National/provincial**: At the national or provincial level, identifying data may be problematic, but the large volume of data means that impacts can be identified fairly robustly.

Addressing all of these issues in evaluating research is made easier by understanding the causal pathways to research impacts shown in the logic model. However, it is not enough to simply identify the pathways: it is also necessary to collect data for indicators that can show progression along those causal paths. This requires that specific evaluation and data collection methods to be employed.

3.4. Evaluation Methods

Frameworks can tell you what data to collect and where to look for it, but they cannot stipulate how to collect it. There are numerous data collection methods, but most fall into two categories: quantitative methods and qualitative methods.

Quantitative methods rely on numerical data. The main quantitative methods used in health research evaluation have been bibliometrics, quantitative surveys, economic analyses, and quantitative scoring of research through expert analysis. Bibliometric analysis is widely used in health research evaluation and is a well understood tool. It uses counts of publications and citations for research publications. This allows for analysis of the quality of research (UK Evaluation Forum 2006; UNESCO Institute for Statistics 2005), identification of world trends in research, and can also apply to patents (technometrics). The use of bibliometrics comes with certain caveats, however: it currently relies on

journal publications,[13] it may not cope well with electronic publications, and it requires expert analysis (Moed 2005a).

Quantitative surveys use closed questions to collect information from researchers about outputs and outcomes, and is a technique that is currently being developed for ongoing evaluation in the U.K. by the Arthritis Research Campaign (arc) (Wooding 2008).

Economic analyses are commonplace in health research evaluations, and use the monetization of health-related impacts (such as improved productivity or the value of health improvement) or the value of commercialized products (Buxton, Hanney, et al. 2004; Funding First 2000; Access Economics Australia Economic Consulting 2003; Murphy and Topel 1999). Studies that try to place value on human health (either through a general value for being healthier or through the improved productivity function) use a number of assumptions to determine the "return" to research, making them difficult to interpret and compare, but easy to criticize. This has been critically assessed by Shiel and Di Ruggerio in Appendix A, p. A43). In the end, however, economic values of research impact can be a very powerful tool for advocacy or political accountability, as evidenced by the response to findings of large returns to health research funding by the Funding First study.

Quantitative scoring of research is often used for research applications, but has also been applied to evaluating research impacts (Oortwijn, Hanney, et al. 2008; Council for Medical Sciences 2002; SciQuest n.d.). Expert scoring has problems in normalizing scores across reviewers and weighting different impacts, however, so it is still rarely used.

Qualitative methods collect information that cannot be counted, providing rich contextual detail. Qualitative methods are particularly useful for explaining how impacts came about, as well as what impacts occurred, however, the findings are often difficult to generalize and compare across evaluations (or potentially even within evaluations). Three main qualitative methods have been used in health research evaluations: case studies, peer review, and open answer surveys (UK Evaluation Forum 2006).

Case studies provide a narrative that allows an understanding of all the factors involved in impacts and can provide a large amount of data for any evaluation. They are research intensive, however, and must follow a standardized protocol in order to be comparable. Peer review is most commonly used to evaluate research funding applications and research outputs (publications), but has also been involved in many evaluations (Dussault, Davis, et al. 2007; Scarpa 2007; National Academy of Sciences Committee on Science, Engineering and Public Policy 1999). Peer review can be very powerful, but is very expensive and time consuming. Open answer surveys also provide qualitative information that can be used in evaluations, but findings can be difficult to interpret.

For any evaluation, a number of different methods must be used in order to triangulate evaluation findings and cover the full range of impacts that health R&D can have (Ruegg and Feller 2003). These methods may be all quantitative, all qualitative, or a combination of both, but should be tailored to collect information that supports the evaluation goals.

[13] With the recent addition of the Scopus database to bibliometric tools, it is likely that it will be possible to analyze books and other literature outside of journal publications.

Table 3. Available methods for evaluating health research impacts (adapted from UK Evaluation Forum 2006)

Method	Pros	Cons
Bibliometric analysis	• Can indicate volume and 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	• Research fields and disciplines need to be taken into account in all analyses • Analysis complicated by the introduction of electronic publications and open and public access journals • Expensive to collect data and analyze • Only able to investigate peer-review publications
Surveys	• Can identify outputs and outcomes associated with particular pieces of funding/research • Provides 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 is involved and therefore what to "cost" • Difficult to *value* many influences involved • Heavily depend on monetary valuation of non-monetary goods (e.g., quality of life) • Difficult 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)	• Potential 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

As well as different evaluation methods, there are also different types of evaluations. For evaluating health research impacts, the two most interesting types are summative and formative evaluations. Summative evaluations are performed at the end of research funding to identify whether anticipated outcomes have been achieved. Formative evaluations are associated with process improvement and therefore occur during the intervention being evaluated (Organisation for Economic Co-operation and Development 2002). These two forms of evaluation are interesting here because although the research itself is being evaluated in a standard summative evaluation, the research funding is being evaluated formatively, since the aim is to improve the process of research funding. This links back to the reasons for evaluation, since summative evaluations lend themselves to accountability and advocacy, while formative evaluations link to evaluation for learning. The findings, methods, and indicators of either of these evaluation types also have a formative effect on future research, since researchers align their goals with those being evaluated. This shows the importance of selecting the right indicators (Chapter 5).

3.5. Data Collection

Large amounts of data could be collected with the available evaluation methods, but it needs to be done in a standardized fashion to allow comparisons of research impacts across funders and, ideally, to research funders outside Canada. Standardizing the definitions for disease groups and types of research is the first step toward standardizing data collection. In the U.K., the U.K. Clinical Research Collaboration[14] has done exactly this by creating a health research classification system (HRCS) that incorporates research activity codes and health categories to provide a matrix within which all research can be classified (UK Clinical Research Collaboration 2007). The system has been successful in the U.K. and is being taken on by a number of other countries to help classify their research endeavours. In Canada, research activity codes must capture all four pillars of health research, so we could not adopt the U.K. system exactly, but adopting a data system that includes some additional elements that are compatible with the HRCS would allow all funders to compare research nationally (and potentially internationally). This aligns with the paper by Hage (Appendix A, p. A79), which argues that meso-level factors—those at the health category level—are vital in understanding the impacts and pathways to impacts of health research.

There is already a large amount of data collected by research funders (mainly on inputs to research such as funding) and by organizations such as Statistics Canada and CIHI. As previously mentioned, using data that has already been collected eases the financial burden of evaluation, but it also reduces the workload on the data provider, making evaluation a less unattractive process for researchers and funders. The data already collected, however, may not be sufficient to provide all the information an evaluator might want, with significant gaps in data collection on secondary outputs and outcomes.

To address these gaps, there are two simple data collection techniques in use elsewhere. First, the use of end-of-grant reporting for researchers is standard for the Social Science and Humanities Research Council in Canada, and is commonplace in other countries for health researchers. The standard form for end-of-grant reporting is a qualitative description of the research findings, process, and likely outcomes, which can be difficult for evaluators to interpret. The Arthritis Research Campaign (arc) in the U.K. has moved away from this to a payback model based survey that asks about common outputs and outcomes from research, with opportunities to add new outputs and outcomes that are not covered. This process makes reporting quicker for researchers, easier to analyze for evaluators, and lets funding be easily compared. It also allows researchers to fill in surveys quickly after a longer time lag, for example, the way arc used a five-year gap after the end of the research to trace outcomes that take longer to come to light (Wooding 2008).

The second technique builds on the common CV already in place in Canada for health researchers (The Common CV System 2006). By expanding the common CV to include a variety of standardized outputs from researchers (such as presentations to public audiences or consultations to government), collecting routine data on researchers' direct impacts will be a simplified process.

Both of these techniques place an additional burden on researchers, and so need to be incentivized. With end-of-grant reporting, the common incentive is to withhold future funding from the researcher until an end-of-grant report is submitted. With the arc end-of-grant survey, there is an additional

[14] The UKCRC is a collection of all the major clinical research funders in the UK, including government, charity, and industry funders.

incentive, because submitting the survey is a considerably smaller burden on the researcher than a full written report (Wooding 2008). Incentivizing a fuller common CV should also be linked to future funding, with all research funders requiring the common CV format to be submitted in any funding application. To make updating the CV simpler for researchers, there could be quarterly reminders from the database requesting updates of publications, presentations and consultations, etc.

4. Chapter 4: Choosing Sets of Indicators and Metrics

4.1. Overview of Indicators and Metrics and Their Use

As identified in the previous chapter, indicators and metrics are different concepts. *Indicators* provide an "indication" of the impacts of an intervention; *metrics* are measurements of the impact itself. As such, indicators can be seen as flags that identify whether the system under evaluation (in this case the health research system) is performing in the way it is meant to. Metrics provide additional information about the specific impacts of a system, however, the more generic the evaluation, the more difficult they are to identify, as they must then relate to all the different types of impact that can occur from the research funded. At a system-wide level, identifying metrics is practically impossible, since there are so many potential impacts that health research can have, and many cannot be quantified across all five impact categories. Therefore, indicators provide a useful, more general tool to address any aspect of health research.

Evaluation methods and data collection are only useful for evaluation if the correct data are being collected, so it is critical to identify the most appropriate indicators of impacts for the research. Indicators can be defined as factors or variables that provide simple and reliable means to measure impacts, changes to an intervention, or performance (Development Assistance Committee Working Party on Aid Evaluation 2002).

Indicators need to be used strategically when performing any evaluation, which means taking into account a number of factors. First, indicators are designed to answer specific evaluation questions, and an indicator for one question is unlikely to adequately answer a different question. For example, if an evaluator wishes to identify the impacts of a research program on health policy research, the most appropriate indicators will be around how policy makers have used research findings. In an evaluation of genomics funding that aims to provide research for industry, indicators for policy impacts are less relevant, and those around industry uptake are more important.

Second, as mentioned earlier, indicators that work at one level of aggregation may not work at other levels. This is highlighted by citation indicators, which rely upon a threshold level of publications (>50) to be statistically viable as a measure of research quality (Moed 2005a). Therefore, citation indicators are not useful for individuals who are being evaluated (and can, in fact, be misleading).

Third, indicators by themselves provide little power for an evaluation, since they can only address single aspects of research impacts. For any evaluation, there needs to be multiple indicators that can identify the range of impacts that can accrue from health research. To use a metaphor from health care, when visiting a doctor for a check up, it would be inappropriate for the doctor to measure only your blood pressure and consider the job done. To stretch the metaphor, if you were being examined for a specific condition, then a single symptom is unlikely to be enough to persuade a doctor to make a diagnosis. In the same way, single indicators can inform a "diagnosis" of research impacts, but need to be backed up by other indicators that can strengthen the diagnosis. This also links to the concept of "ethically inclusive use of ROI," outlined in the MacDonald and Knoppers paper (Appendix A, p. A121), which suggests that using multiple indicators to identify impacts can best incorporate an ethical aspect to evaluation, since it spreads the ethical burden across multiple indicators.

The three strategies identified above all relate to summative evaluations, but there are also issues to take into account when considering formative evaluations. As mentioned, formative evaluations can

shape the way an intervention develops, meaning that indicators can drive behaviour within the intervention.[15] For health research, this means that the indicators chosen for a formative evaluation will be seen as goals or targets for researchers. This is also an issue for summative evaluations that are part of a standard reporting process, since future research will tend towards the indicators chosen.

Indicators fall into the specific impact categories identified in the framework: advancing knowledge, building capacity, informing decision making, health impacts, and broad economic and social impacts. Within each of these categories, indicators have been used worldwide to determine impact, and it has been the job of this panel to showcase which are most appropriate and why.

4.2. Defining "Appropriate Indicators"

Identifying indicators for evaluating health research impact is a major challenge. The framework developed can help to define indicators, however, since it identifies the impacts and pathways to be evaluated, and thus both where and what types of indicators are required. Additional criteria must be used to select the most appropriate indicators for most impacts. These criteria relate to individual indicators (showcasing their strength as an indicator), and also to groups of indicators (highlighting the way groups can work together to produce informative, robust, and ethically sound evaluations).

Individual indicators must adhere to criteria that can be split to showcase the attractiveness of indicators (that is, how attractive are they for an evaluator to use) and the feasibility of indicators (or how likely is that they could be used) (Butler 2008). This is the approach being taken forward in Australia, where indicators have formed part of the Research Quality Framework (RQF) for distributing research funding to Australian universities for a number of years (Research Quality Framework 2006).

Box 2 shows the attractiveness and feasibility characteristics for individual indicators, based on work on indicators from Australia and the U.K. (Butler 2008). By adhering to the criteria for feasibility and attractiveness, individual indicators can be considered as "appropriate," although as previously mentioned, appropriateness is related to the evaluation questions being asked and the aggregation levels being investigated.

[15] As an interesting aside, improved data from formative evaluations can help with an evidence-based partition of funding between different types of research (be that different pillars, health conditions, or public/private levels of funding).

Attractiveness:

Validity – does it relate directly to a critical aspect of the research?

Behavioural impact – does it drive behaviour in a particular direction? Is it likely to result in any negative, unintended consequences? Does it create "perverse incentives?"

Simplicity – is the methodology, and the strengths and weaknesses relating to the indicator, readily apparent?

Coverage – does it cover a large proportion of output for the fields of research to be assessed?

Recency – do the data relate to current research performance, or look over a longer timescale?

Methodological soundness – is the calculation of the indicator methodologically sound and statistically robust?

Replicability – can the indicators be used year on year in a comparable fashion?

Comparable – do other organizations collect comparable information or have targets to benchmark against?

Relevant – is it relevant to what the organization is aiming to achieve (linked to evaluation questions for that organization)?

Responsiveness – is the data sensitive/responsive to the Minimally Important Difference (MID)?

Feasibility:

Data availability - do the data needed to derive indicators exist, and do both the analysts and those being assessed have access to it?

Cost of data – how expensive is it to purchase the data on license?

Compliance costs – how labour intensive is it to extract/obtain the data?

Transparency – can the calculations be replicated by interested external parties?

Timeliness – can the data be obtained/provided relatively quickly?

Attribution – can the data be discretely ascribed to the unit being assessed? Direct attribution is ideal, but unlikely; using attribution as a concept is important, though, as it provides a link between the impact seen and the research.

Avoids gamesmanship – does the indicator provide scope for special interest groups or individuals to game the system?

Interpretation - can the data be open to misinterpretation or misuse by commentators and/or actors using the evaluation findings (for example, university league table rankings)?

Well-defined – does it have a clear, unambiguous definition so that data will be collected consistently, and so that the measure is easy to understand and use?

Box 2. Criteria for appropriate individual indicators

Attractiveness and feasibility are not distinct concepts and there are trade-offs between the two. For example, having a methodologically sound and wide coverage indicator (attractiveness criteria) is rarely cheap and therefore conflicts with the compliance costs and costs of data collection (feasibility criteria). Balancing the attractiveness and feasibility of indicators is an important part of the evaluator' role, and can be made easier by considering groups of indicators for evaluations.

As a subset of the validity of indicators, there is a need to understand the appropriateness of an indicator for the particular aspect of research it is evaluating (such as the development stage in pharmaceutical development), providing appropriate expectations of the likely outcomes from that stage. For example, it would not be appropriate to use sales of drugs as an indicator for the first stage of pharmaceutical research, but it would be appropriate at later stages of development and marketing. This is an example of using an indicator that does not relate to the critical outputs of the research being evaluated (the definition of validity), even though the indicator may be sound for other stages of the research process.

Any good evaluation requires an understanding of the logic of the research impact process (through a framework, for example), and a balanced view of the impacts arising from the research under investigation. Since there is no single indicator that can provide this kind of evaluation, it is necessary to produce a suite of indicators (converging partial indicators) that can showcase the impacts and triangulate findings (Martin 1996).[16]

Box 3 shows the FABRIC criteria that need to be met for groups of indicators (HM Treasury, Cabinet Office, et al. 2001).

Focused on the organization's aims and objectives
Appropriate to, and useful for, the stakeholders who are likely to use it
Balanced, giving a picture of what the organization is doing, covering all significant areas of work
Robust in order to withstand organizational changes or individuals leaving
Integrated into the organization, being part of the business planning and management processes
Cost effective, balancing the benefits of the information against the costs

Box 3. FABRIC criteria for "appropriate" groups of indicators (HM Treasury, Cabinet Office, et al. 2001)

Many of the factors that apply to individual indicators also apply at the group level, since a group of indicators used for an evaluation must also have a wide coverage (similar to the *balance* criteria) and not drive behaviour in inappropriate directions. This second concept is important, since it is possible to create a suite of indicators that can drive the behaviour of researchers in desired directions, even if

[16] This links with the concept of "multifactoral comparisons" identified by MacDonald and Knoppers in Appendix E, where they argue that for ethical evaluation, "complexity is essential, since ethical choices generally involve context-sensitive weighing of multiple factors."

individual indicators would normally drive behaviour in undesirable directions. For example, while research quality indicators have focused on citation data (driving researchers to focus on peer-reviewed journal publications), adding measures of broader dissemination (such as presentations to the general public and consultancy to policy makers) to the suite of indicators can drive researchers away from only publishing in journals.

By taking these factors into account when determining indicators and groups of indicators, there is a validation process that is open and accountable (linking to the need for ethically sound evaluations and evaluations that employ "ethically inclusive use of ROI" – MacDonald and Knoppers, Appendix A, p. A121).

Indicators provide data on the outputs and outcomes of research, and therefore bring problems of privacy and sharing of data, as alluded to in Chapter 3. Privacy is an issue for individuals in several health indicators, since they rely on health data measured by the health system. For example, one of the best ways to access changes in the health of individuals and their movement through the health system is through electronic health records, yet the privacy constraints for electronic health records can be very strict with regard to the use of information. Health data privacy is not an insurmountable issue for improving health research, as shown by the sale of population genomic data in Iceland for medical research purposes (Genome vikings. Editorial. 1998).

Sharing of data is a slightly different issue, and one that resonates particularly for pan-Canadian evaluation. With many different funders and researchers involved across the country and across different funding sectors, there are issues around proprietary information that may make organizations reluctant to share. Sharing information to produce Canadian analyses also requires there to be a central repository for the data, and an organization designated to collate and control that data. This is an additional financial burden on whichever group controls the central repository. Incentivizing the collation of data is an important aspect of facilitating data sharing.

4.3. Identifying Appropriate Indicators

Taking into account the factors mentioned above, we have identified a number of appropriate indicators within each impact category of the evaluation framework described in Chapter 2. In addition to appropriate indicators that can be put in place now, there are also *aspirational indicators* that would capture additional information on knowledge impacts. They are aspirational because they cover data that are highly desirable, but which are currently difficult to collect and/or analyze. For each indicator, we also identify the level at which the indicator is appropriate (for example, individual researchers, research groups, institutions, etc.) and the pillars of research for which it is most relevant.

4.3.a. Advancing Knowledge

Most work on research indicators has focused on the "advancing knowledge" category, since it deals with the outputs of research that represent the aspect for which researchers and research funders are directly responsible, and the impacts with the most straightforward attribution to research (Wells and Whitworth 2007). There are four subcategories for advancing knowledge: quality measures, activity measures, outreach measures, and structural measures.

Quality Indicators

High quality research is an expectation of research funders. For example, the CIHR mandate requires research excellence (Canadian Institutes of Health Research 2007d). Quality can be measured using peer review of publications (an expensive and time consuming process) or through citation analysis, which incorporates peer judgement by analyzing how often a publication is cited. Citation analysis can be performed relatively easily, and recent work suggests that both citation analysis and peer review arrive at very similar quality judgements (Van Raan 2006).

Relative citation impact: The relative citation impact indicator is the average number of citations received by the unit being analyzed (be that a group, department, institution, or even Canada as a whole), compared to the world citation rate for the disciplines being investigated. This measure must use discipline-specific benchmarks to account for different citation practices across disciplines, and is only be considered robust if based on a sufficiently large set of publications (usually >50 publications) (Moed 2005a). Individual researchers (and junior researchers, in particular) generally produce too few papers for robust analysis. Relative citation impact can be applied to any of the four pillars (or even interdisciplinary research), since it accounts for the discipline of health research being assessed. For pillars III and IV, however, there are problems with the proportion of publications represented in citation databases.

Highly cited publications: The highly cited publication (HCP) count indicator represents the number of publications by the unit that are in the top *X* percent of cited world publications in the discipline. It commonly uses the top 10%, 5%, or 1% of the world's publications in a discipline. It provides a measure of the number of the highest quality research publications produced, and complements the average quality measure of the relative citation impact. Since it measures the number of top publications, HCP counts do not rely on a minimum number of publications, and can be used for individuals as well as larger aggregations of researchers. As with all current citation measures, HCPs are less reliable for determining quality in pillars III and IV, since these two pillars publish outside of indexed journals more regularly than pillars I and II. Relative citation impact and HCPs are the two most reliable citation measures currently available (Moed 2005a).

Publications in high quality outlets: Counting the proportion of publications (journals, publishers, conferences) that appear in outlets judged to be of high quality can help to account for pillars III and IV, where a smaller proportion of knowledge production is in the journal literature. Judging high quality outlets is being undertaken in a number of countries (for example, the "European Science Foundation Humanities" project, and the "Excellence in Research for Australia" journal and publisher rankings), identifying discipline-specific top outlets. As a proportion of publications, this can be applied to individuals and higher aggregations. It should be noted, however, that if most of the publications are appearing in indexed journals, then citation analysis should be used, since citations focus on the specific publication, not the average for publications in an outlet. The publications in high quality outlets indicator can also be modified to investigate publications in outlets that target a specific audience, such as those used by health practitioners.

Activity Indicators

Activity indicators showcase the amount of research output for a given unit. They are used to show productivity of individuals and units, and are sometimes falsely used to imply quality. (Caution: It is vital that activity indicators are not used without accompanying quality indicators, because increasing

activity alone is a risk to quality and can lead to perverse incentives for researchers to prioritize volume over quality.)

Share of publications: The share of publications indicator is the output or number of publications of the unit under study as a proportion of a reference output, usually the level of aggregation above the unit under study (for example, for a university department it would be for the university as a whole; for a funding agency it would be national output). This should normally be used within a discipline rather than for total publications, since different disciplines have different publication practices (Moed 2005a). Currently, share of publications can be easily calculated for indexed journal articles, but not for other types of publications. This is because it may be difficult to obtain data on the reference output (for example, how many books are produced by Canadian researchers in nursing?). The share of publications can work for any aggregation above an individual, but not for individuals, since their output is usually too small. Since share of publications has the potential to take into account all publications, it can be applied to all pillars of health research. To include a quality measure in this indicator, it could be calculated for share of HCPs only.

Publication counts: Counting publications is the simplest way to measure output, and can be done easily and cheaply (the data are routinely collected in order to calculate other indicators such as publication share and relative citation impact). By themselves, however, counts are a meaningless indicator, as they require a comparative aspect (Butler 2002). For example, is the level of output above or below what is expected in that discipline? Are the papers of any significance? Are they cited by others? Publication counts alone should only be used where shares of publications cannot be calculated.

Outreach Indicators

Advancing knowledge also requires that research findings are distributed to other researchers, to allow further research. The emphasis on this outreach is currently on interdisciplinary research, both in terms of conducting the research and in the use of its findings.

Co-author analysis: This indicator uses publication author information to identify the proportion of publications that are co-authored: internationally, nationally, with industry, with other disciplines, etc. (Glanzel and Schubert 2004). Determining which co-authors to investigate depends upon the focus of the evaluation undertaken. For example, one could use interdisciplinary co-authorship and co-authorship with industry in evaluating cross-disciplinary funding such as the “Networks of Centres of Excellence” (NCE) program (R.A. Malatest and Associates Ltd. and Circum Network Inc. 2007). Since this indicator investigates all publications, it can be used for individuals as well as groups and larger aggregations. It can also apply to any pillar, although because it may be difficult to obtain address details of non-indexed publications, pillars III and IV may suffer from a smaller percentage of their output being usable for analysis.

Field analysis of citations: While co-authorship analysis may indicate whether research is conducted with other disciplines, field analysis of citations examines whether the findings are taken up by different disciplines. This indicator uses the proportion of citations that come from fields outside of the field of the publication. As a citation proportion, it can be used for individuals as long as they have a large enough publication record (>50 publications), but is more useful for groups and larger aggregations. Since it uses citation databases, this indicator is currently most useful in pillars I and II, but can be applied to III and IV with a caveat about the coverage of the analysis.

Contextual/Structural and Indicators

It is important to understand the breadth of research activity for an organization, since this information allows an analysis of strengths and weaknesses, and identifies the breadth of the research portfolio of the unit. This structural analysis of research outputs places an emphasis on the extent of activity in specific areas of health research.

Relative activity index: An indicator used by the Center for Science and Technology Studies in Switzerland (Research Evaluation and Policy Project 2005; Rehn, Kronman, et al. 2007), the relative activity index (RAI) identifies the proportion of an organization's output that is linked to specific disciplines, thus highlighting the main research focus of the organization. By using only HCPs in the creation of a relative activity index, this indicator can showcase the research focus, while still accounting for quality (avoiding a perverse incentive to produce a large volume of lower quality publications—an incentive that we tacitly recommend in our "Activity" section). Since this requires a large publication output, and since by their very nature smaller units have a relatively narrow focus, it is most applicable at an organizational level or above. As with the share of publications, this indicator requires a benchmark at the level of aggregation above that analyzed (for example, for national RAI, the benchmark would be the world publication balance in health research). As a ratio of publications, this can be applied to all pillars of research, but as with many indicators, it is currently more efficacious for articles that can be readily classified to disciplines on the basis of the journals that carry them.

Aspirational Indicators

The indicators mentioned above are appropriate for now, as they represent data that can be collected and indicators that have been validated. As noted, however, most current indicators tend to work better for pillars I and II than for pillars III and IV, because they rely on publications indexed in databases as the major form of output. Ideally, we would also like to access other forms of research output, such as books, reports, and articles in non-indexed journals. Also, as scientific publishing moves away from paper journals to electronic ones, we would like a way to track downloads as an equivalent measure to citations (since this would allow direct comparison of journal and non-journal publications).

Expanded relative citation impact: This indicator expands citation analysis to cover a greater range of publications, including book-to-book citations. Work is ongoing in this area, and with the advent in recent years of on-line tools such as Google Scholar as alternatives to Web of Science and Scopus, this area of citation analysis will continue to grow. As with citation analysis for journal publications, this indicator still relies on a minimum number of publications in order to be statistically accurate, so it is not appropriate for individuals. This could prove especially important for pillars III and IV, where a greater proportion of output is in non-journal literature.

Relative download rate: The relative download rate indicator is the average number of downloads per publication when compared to a discipline benchmark (since download activity is likely to be different for different disciplines). Ideally, downloads would be able to differentiate between the audience downloading (for example, government versus public downloads), and would control for web-crawling (a serious problem in measuring the downloads from many databases). Downloads could be weighted so that they could be directly compared to citations (Moed 2005b). It may also be possible to create a "most downloaded" equivalent of the HCP measure for citations once data on downloads has been collected. Having a relative download rate and a highly downloaded publications (HDP) measure would

allow non-journal publications to be analyzed in the same way as journal publications, making downloads a particularly useful measure for pillars III and IV. It should be noted that downloads (like citations) are only recommended for groups, organizations, or larger aggregations, and not for individuals.

This indicator has the potential to cover any publication in repositories, including books, book chapters, reports, etc. With the recent move by NIH in the USA to insist that all publications resulting from their funding be deposited in Pubmed within six months (National Institutes of Health 2008), and that CIHR publications be open access within six months (Canadian Institutes of Health Research 2007c), the likelihood of this indicator becoming feasible has increased dramatically. It is notable that the U.K. (and potentially Australia) also seems likely to make depositing publications into Pubmed (or a similar database) a requirement (Terry 2005).

4.3.b. Capacity Building

Capacity building is an important aspect of many research funders' activities, and indicators fall into subgroups that represent personnel (including aspirational indicators for improving receptor and absorptive capacity), additional research activity funding, and infrastructure.

Personnel Indicators

Personnel capacity measures the improvements in staffing or the capabilities of people to do their jobs (or future jobs).

Graduated research students in health-related subjects: Measuring the numbers of graduated PhD, MSc or MD students in health-related disciplines identifies the possibility to improve the future research capacity of Canada (part of many funders' remits). Statistics Canada collects data on graduates with PhDs (King 2008), and this could be disaggregated to show health-related PhDs (as well as student gender, subject areas, universities, etc.). Additional data on MSc and MD degrees could be collected in the same way. As a monitor of new research degrees, this indicator is not appropriate to evaluate individuals in receipt of research funding (the students themselves), but it is useful for evaluating institutions such as universities. It is most useful for evaluating funders of research students, and at the provincial and national level, against benchmarks in other fields of research and other countries (and year on year). This indicator is applicable at all levels. As an aspiration, this indicator would expand to track graduates through their career progressions, using a tool such as the Statistics Canada National Graduate Survey (Statistics Canada 1999). This tool could also provide an indication of the quality of individual or higher-level training environments.

Number of research and research-related staff in Canada: Statistics Canada already collects information on all research staff (researchers, technicians, and support staff) across all types of research organization (Statistics Canada 2008b), but does not limit it to health researchers. Since the data comes from surveys of research organizations, it would be possible to request classification of health researchers by pillar. In its current state, this information is only really suited to pillars I and II, since it identifies "natural sciences" researchers rather than health researchers. However, if data are available for all research staff, then it could apply to all pillars. As with graduate students, this indicator does not provide information about individual, funded researchers, but is useful to identify how well institutions, provinces, and Canada as a whole are maintaining and building a human capacity for health research. This would need to be benchmarked year on year, and compared to capacity in other research areas.

Funding Indicators

Funding allocated to research is a well understood system at all levels, from funders (Canadian Institutes of Health Research 2007e) to the national situation through GERD, GBAORD, and BERD as proportions of R&D (Fast 2007).[17] However, research outputs can have a significant impact on the way research funds are allocated in the future, so it is useful to know where research has been a factor in facilitating an increase in research activity funding capacity.

Levels of additional research funding: Identifying the additional funds that are attracted by research activity is a useful way to identify activity funding capacity building, and compliments the standard year-on-year total funding that organizations provide. It is measured as funding from "external" sources that can be attributed to the capacity built in an organization, institution, or region, and would include matched funding. This has been used by provincial funders to identify national funds pulled into the province by the research capacity in place (Birdsell and Asselbergs 2006). This indicator is only recommended for funders, provinces, and nationally, since it is only useful for levels that provide funding, and because it is often impossible to disaggregate the role of the individual within a group, particularly in relation to grant income (Gläser, Spurling, et al. 2004). The main disadvantage of this indicator is that it is difficult to attribute additional funds to research that has been funded by a specific province/organization, since researchers tend to receive funding from multiple funding bodies. This can lead to a risk of double counting additional funding. This indicator can be used for all the research pillars.

Infrastructure Indicators

Performing health research requires infrastructure as well as activity funds. Improvements in infrastructure often come down to research findings that indicate the need for further investigation given improved infrastructure. For our purposes, infrastructure can be new research equipment, databases for capturing data, new laboratories, or new buildings.

Infrastructure grant values: This indicator is the amount in dollars of infrastructure funding pulled in by a research project, group, or organization. By measuring funding rather than infrastructure itself, we can capture the different aspects of infrastructure (equipment, databases, buildings), since they all come from infrastructure grants. Because there is limited infrastructure funding, the value pulled down represents the importance of that infrastructure in the eyes of those controlling funding. The main issue with this indicator is that it misses out on infrastructure from other sources (for example, university re-allocation of space, etc.). As a measure of acquired infrastructure funding it is only recommended for institutions, organizations, provincially, and nationally, since it is at these levels that infrastructure is primarily acquired. Although infrastructure costs differ for each pillar of research, and for different projects within the pillars, this indicator can be used in any pillar. It is less useful for comparing across pillars.

[17] It is worth noting that the use of input data in evaluation is only useful if it is clear that there is a "right" level of funding that is not being attained. Currently, there is very little information about what correct funding levels are or should be for different disciplines and countries, meaning that funding is important to track, but does not provide the best information on how to improve impacts from any form of R&D. Collecting input data is also valuable if an evaluator wishes to perform any sort of cost-benefit analysis.

Percentage of activity grants with infrastructure support: Infrastructure alone is of no use, and funding should be tied to research activity. By measuring the co-ordination of infrastructure grants with activity grants, we can identify whether the infrastructure and activity are occurring together to allow the most efficient production of research results. This measure does not account for research that has no new infrastructure costs, or for research that is covered by universities, but it is useful for tracking allocations of infrastructure funds. As with the value of infrastructure grants, this is only recommended for institutions, organizations, provincially, and nationally. Accessing data for this indicator requires the collation of activity grants awarded by organization/group, and then identifying linked infrastructure grants—meaning that the activity and infrastructure funders need to collaborate more closely in data sharing. As a measure of the match-up between activity and infrastructure funding, this is applicable to all pillars and, as a ratio, can be useful in comparing across pillars.

Aspirational Indicators

Receptor capacity: This indicator represents the ability of those in policy and administrative positions to take research findings on board, since without the ability to identify the best research, evidence-informed policy is very unlikely. Receptor capacity is generally only measured to evaluate specific learning initiatives for policy makers (for example, Denis, Lomas, et al. 2008), but the surveys used for such processes could be modified to survey policy makers and administrators more generally. Receptor capacity can relate to any pillar of research, but is most likely to be useful for research from pillars III and IV (Canadian Health Services Research Foundation 2007; McDaid and Cookson 2003). Surveying policy makers and administrators could lead to the identification of specific research findings, but it is unlikely that policy makers can pinpoint the studies that have informed their thinking (Wooding, Nason, et al. 2007). This indicator is most appropriate for identifying whether research in general is getting through to policy makers and administrators. The main drawback for this indicator is the likely response rate to surveys. Incentives would be necessary for policy makers and administrators to be involved (such as relating research to their performance appraisal).

Absorptive capacity: This indicator represents the ability of those in research to adopt the research findings of others and exploit that knowledge. Absorptive capacity is most commonly attributed through collaborations (particularly collaborations between industry and academia), R&D funding intensity, and co-authoring (Cockburn and Henderson 2003; Griffith, Redding, et al. 2003; Schmidt 2005). It would also be possible to track the disciplines cited by researchers in publications as an example of absorptive capacity across disciplines, although this would not highlight research in the same discipline coming from other sources. Of these, no single measure adequately captures the concept of absorptive capacity, but since R&D funding intensity is collected for firms, and collaboration and co-publication indicators feature elsewhere (in “informing decision making” and “advancing knowledge” indicators, respectively), capturing multiple measures should not come at an additional cost to any evaluation. This measure could be used for organizations, provinces, or nationally to measure absorptive capacity, but is not relevant for individuals or groups because it relies on higher level collaboration and funding data. As a measure of the ability to take the research of others on board, it can apply to all pillars of health research. Measuring the absorptive capacity across disciplines is a particularly useful indicator for research that crosses the four pillars.

4.3.c. Informing Decision Making

Informing decision making indicators are problematic, because they represent the pathways from research to its outcomes on health, wealth, and well-being. They can be thought of as falling into

health-related decision making (where health is broadly defined to include health care, public health, social care, and other health-related decisions such as environmental health); research decision making (or how future health research is directed); health products industry decision making; and the general public's decision making (an issue covered in more detail by Picard in Appendix A, p. A138). These decisions can be made at a variety of levels from individuals in the general public or health sector, to national policy decisions or organizational policies.

It is important to remember that the decision-making process for any grouping is often the result of a number of competing factors, of which research findings is only one. It is also worth noting that it is often difficult to determine which factors are involved in making a decision. For these reasons, measures of research impact on decision making are often proxies that involve research informing a group, but not necessarily directly influencing the decision-making process.

Health-related Indicators

Health research should impact on the decisions of all health professionals and other decision makers in the health system. This includes decisions in all aspects of the health system (from primary and secondary health care, through to public health and social care), those areas of policy and industry that relate to health (such as safety at work), and education for the health sector (for new health practitioners and ongoing education for current practitioners). Within the four subcategories that represent the different aspects of a broad health system (health care, public health, social care, and other health-related systems), there is a three-layer hierarchy of data sources for informing decision making indicators. The top level involves published evidence that identifies research; the middle level, surveying decision makers to identify what has influenced them; and the bottom level, asking researchers to report on how their research has informed decisions. The "most appropriate" indicators identified here are based on the most likely available information for each aspect of health-related decision making (so, if higher levels of information are not readily available, we recommend collecting information at the level below).

HEALTH CARE

Use of research in guidelines: By analyzing citations to research in clinical and service guidelines, we can start to understand which research is going on to have an effect on practice. Although this measure is imperfect (as it captures only the formal routes through which research reaches clinicians, not informal routes such as conversations between clinicians), it has been used in other research into the impacts of health research (Buxton, Hanney, et al. 2008; Grant, Cottrell, et al. 2000) and is something that CIHR is are committed to analyzing (Canadian Institutes of Health Research 2005b). This indicator can help to identify specific research informing health care and the proportion of Canadian research informing health care (Grant, Cottrell, et al. 2000). It is more problematic for linking research to funders, as there is no current requirement to identify the funder in the publication. Guideline analysis could be applied for individual researchers, but as a citation measure would result in very small numbers for individuals. It is more practical at aggregate levels (group/institution/ province/nation). Since citations in clinical guidelines are predominantly to clinical research publications (Grant, Cottrell, et al. 2000), this indicator is most appropriate in pillars II and III, but could be applied in pillar I using generational analysis of reviews cited by the guideline.

PUBLIC HEALTH

Survey of public health policy makers: Public health decision making does not follow the same guideline culture as medical care. To access the research making an impact on public health decision making, one must survey public health policy makers about what research they have used to inform their policies. Surveying this group may be difficult unless their time is incentivized, perhaps through evidence-informed policy being built into performance appraisals. This indicator is unlikely to be useful for individuals (since policy makers may not know the individuals behind research results), but it may be useful for groups. Although it is mainly applicable to pillar IV, any health research could conceivably inform public health policy (for example, all four pillars have informed the smoking bans in place in Canada).

SOCIAL CARE

Use of research in guidelines: As with health care, social care can be informed by research that is used to create social care service guidelines. Analyzing these citations is performed in the same way as for clinical guidelines, and comes with the same caveats: this indicator is not appropriate for individuals, but is more useful higher aggregate levels (group/institution/province/nation). It is most likely to be useful for pillars II, III, and IV, since it is unlikely that basic biomedical research will inform social care.

OTHER

Researcher reported use of findings outside health: Since the impacts of health research outside of health can be very wide-ranging, single indicators are hard to identify. As such, it is necessary to ask researchers to report on those areas outside of health, public health, and social care where they have had an influence. For example, health research findings could be picked up by transport or employment policy to improve safety or working conditions. Since this indicator would use information from individual researchers, it could be applied to individuals, but since the number of impacts for any individual is likely to be small, it would be better used at institution/funder levels. The problem with this indicator is that researchers may not know if their research is used outside their area of research. This indicator could be applied to all pillars.

HEALTH-RELATED EDUCATION

Research cited in ongoing health professional education material: Educating health practitioners includes training and continuing education. Materials produced for continuing health professional education cite research to support new practices, and these can be analyzed in the same way as guidelines analysis. There may be issues accessing the references for these materials, since there is currently no standard database of educational materials, however, such a resource could be created without a great deal of difficulty. This indicator can be linked to individuals through their citations, but as with guideline analysis, this is likely to produce small numbers. As such, this indicator is more appropriate at group, institution, funder, provincial, and national levels. Since the data collected would be relevant to any health practitioner, this indicator can apply to all four pillars. Note that early health professional education is covered later (in the “research education” indicator section), since it uses data that is also applicable to training for health professionals.

Research Indicators

Research findings inform future research through informing future funding decisions, changing research policies (at any level from government to institutional), and informing the education of those in research or going into research.

RESEARCH FUNDING

Citation analysis of successful funding applications: This indicator uses cited research in successful funding applications to identify underpinning research that may inform a new research direction. Two issues present with this indicator. First, only research funders themselves can access references in successful applications, meaning data would have to be shared between funders. Second, funding applications tend to contain a lot of self-citations, but citation analysis can take this into account and ignore (or account for) self-citations. As with other citation-based methods, this is not recommended for attributing research funding impacts to individuals, since the number of citations is likely to be small. By aggregating citations to research groups, institutions, and higher aggregations, however, this indicator can provide information on the research driving future research in all four pillars.

RESEARCH POLICY

Consulting to policy: The research policy indicator is the number of consultations to policy makers (where "policy" is broadly defined as organizational to national policy) by researchers. Since there is no "correct" number of consultations to policy, this needs to be tracked in a year-on-year analysis. Data could be collected through policy makers' official documents or through researchers. We recommend surveying researchers, since a top-down approach neglects unofficial consultation. Consultations could be asked about as part of end-of-grant reporting surveys (Wooding 2008). This indicator is useful for individuals, as it can help to identify which individuals are strongly linked to policy circles—a factor in the incorporation of research into policy (Wooding, Hanney, et al. 2004). It can also be aggregated to levels above the individual (groups and institutions). Since there is no desired level of consultation, however, this indicator is less useful at higher aggregations such as the provincial or national level. This indicator is applicable in all pillars.

Requests for research to support policy: Consulting to policy provides information on personal requests for research information, but identifying the requirements of policy makers for research in general requires looking at research documents. By identifying the number of requests for research made by policy makers (primarily systematic reviews of research), we access the desire for research information at policy levels. Since this only determines the level of interest in research, it is not something that research funders can influence directly, but it does provide valuable information for the health research system on the way it is influencing a desire for research findings at the policy level. Data can be collected through official requests for research (for example, systematic reviews commissioned), or through researcher responses to requests (collected, for example, in end-of-grant reporting surveys). As a measure of interest in research, this is not an indicator that can be attributable at the individual, group, or institution level. It is only relevant at a provincial or national scale. Since it collects data on all research requests, it is applicable to all pillars of health research.

RESEARCH EDUCATION

Research used in curricula for new researchers: To identify the research informing education, we can analyze the text books and reading lists used by university students in health-related disciplines for the research publications they cite. Advances in citation databases, such as Scopus' book collections, suggest that the current problem of accessing citations in textbooks could soon be solved. The first step in identifying the data involves universities creating a list of textbooks and recommended reading (something that already exists to inform students) in which to search for citations. Assuming that we can access the citations in all of the textbooks, the first half of this indicator is simply to count how

many times researchers and aggregations above that (up to Canadian publications) are represented. Ideally, we want the best research to be the research that influences teaching. Once the publications are available, it is possible to identify what proportion of those publications are in the world's top 5% highly cited publications (HCPs) for their research field. Once we start identifying the HCPs, this indicator can only apply to the group level and upward, since citation quality measures (as opposed to counts) should not be used at an individual level. As a measure of education for all health research disciplines (including health professional education), this indicator can apply to all pillars.

Health Products Industry Indicators

The health products industry needs access to the best research in order to create the best products. There are a number of routes for accessing that information, from formal licensing of patents through to co-authoring on research. There are also ways to identify which research influences health products industry research as it progresses through stages toward marketing and sales.

Number of patents licensed: Counting licensed patents provides information about research that has been used by industry. This can be benchmarked against previous years or against internationally held patents (Organisation for Economic Co-operation and Development 2007). There are already Canadian groups (such as Science-Metrix) that produce patent analyses for Canada (Science-Metrix 2008), and data on licensed patents is already maintained in Canada and reported on by the Treasury Board (Treasury Board of Canada Secretariat 2007). Licensed patent counts (Treasury Board of Canada Secretariat 2007) can be used for individuals, but are most useful at group, institution, province, and national levels, where sample sizes are larger. Since most patented health research is likely to occur in biomedical or clinical research, this indicator is most appropriate for pillars I and II, and could be very useful for identifying whether cross-pillar research is more successful at producing licensed patents.

Clustering/co-location: Co-location analyses can show where industry is located in relation to academic centres—a factor that facilitates the translation of research between academia and industry (Gunasekara 2006). This analysis, which uses geographic data, can provide an overview of where innovation and knowledge transfer is likely to occur, allowing resources to be focused on likely areas of knowledge transfer. Since this indicator looks at geographic regions, it is only useful at provincial and national levels. It could be used for any of the four pillars, but is most likely to be relevant to pillars I and II.

Consulting to industry: The interactions of researchers and industry can be traced using year-on-year values for consultation. Data can be gathered through company reports or through researchers (as part of an expanded CV or end-of-grant reporting). This indicator could be used for individuals to identify those translating to industry; or for groups, institutions, or provinces, to identify environments that are conducive to knowledge translation. Industry now seeks information from all pillars, but it is less likely that there will be requests for researchers in pillar IV.

Collaboration with industry: Consultancy provides information about questions asked of researchers, but collaboration (as measured through co-author analysis) identifies where researchers are working with industry on research projects. Since this indicator involves bibliometric analysis, it relies on industry publishing research findings in journals. As an indicator that uses publication data, this is not recommended for individuals (since the likely sample size is too small), but would be recommended for groups, institutions, provinces, and nationally. Co-authorship analysis can identify research from any pillar of health research, although it is most likely that the research will come from pillars I and II.

Use of research in stage reports by industry: Research in industry must go through six stages in order to become a successful product: preliminary investigation, detailed investigation, development, validation, commercialization, and sales. Citation analysis of the reports produced between stages (Government of Canada 2008) could identify what research is underpinning the movement of products through the development pipeline. The key issue with this indicator is that it relies on accessing stage reports for industry and mining citations from them. This would require that a group take responsibility for data gathering and analysis. Since this indicator uses citation analysis, it is not recommended for individuals, but only for groups, institutions, provinces, and nationally. The research identified is most likely to be in pillars I and II, but analysis would identify any research used by industry.

General Public Indicators

As noted in the paper by Picard (Appendix A, p. A138), the public is an essential group for health research to connect with, though this is something that can occasionally go awry (Canadian Broadcasting Corporation 2007; Van Driel 2002). Although it is possible to survey members of the general public about how research has influenced their decision making, this is often a difficult question to answer and proxies must be found to address the link of research to decision making. It can also be difficult to estimate how health research influences decisions, as opposed to simply entering the ethical and scientific consciousness of individuals. The public is considered here to encompass groups that have an interest in particular health issues and the public more generally.

ADVOCACY GROUPS

Research cited in advocacy publications: Research that is mentioned 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. The key issue with this indicator is that it misses work for advocacy groups that is not cited (such as consultations), but this information can captured through an expanded researcher CV (The Common CV System 2006) that collects information on researcher activities as well as outputs. As a measure of research cited, this indicator is not recommended for individuals, but is recommended for groups, institutions, provinces, and nationally. It also is recommended for research in all four pillars.

PUBLIC EDUCATION

Number of public lectures given: To identify the direct impact of health research on the public, we can capture the number of lectures given to public audiences. These data are collected through an expanded standard CV (The Common CV System 2006), and are applicable for individuals and any aggregation of data above that. As a measure of the involvement of individuals with the public, this indicator can be applied to all pillars.

Aspirational Indicators

In order to collect data on certain aspects of decision making, we propose two additional “aspirational indicators” that (if data were collected and analyzed) would broaden our understanding of how research goes on to impact on the decision making of individuals in the public and those in policy making. These aspirational indicators represent data that is available, but which cannot be identified robustly because of the lack of an appropriate collection mechanism.

MEDIA

Media citation analysis: The media is a useful proxy for research informing the general public, since it can have a significant impact on the research findings to which the public are exposed. By analyzing mentions of research in newspapers, it would be possible to create a database of the research that is informing the media. Work is currently ongoing to create a potential international database of major national newspapers, which would require individuals to identify research mentions in newspapers on a daily basis. Although this indicator covers only print media at this time, it is possible that it could be expanded to cover web-based media too. Addressing TV and radio would be more problematic, but it is unlikely that research covered by TV or radio would not also be covered by the print media (and if this were expanded to web media, it would almost certainly cover TV and radio, since both have websites for their news sections).[18] Funders currently collect information about their media coverage (Canadian Institutes of Health Research 2005a; UBC Faculty of Medicine 2008), so this indicator could dovetail with current practices. Since the media tends to mention individuals, this indicator is recommended at the individual level and aggregations above. It is also relevant to all pillars.

PUBLIC POLICY USE

Citations in public policy documents: Analyzing citations to research in public policy documents (grey literature) would be a valuable way to identify exactly which research is informing policy. The possibility exists that the development of new bibliometric tools would allow citation analysis of policy documents (Lewison 2004), and that Google Scholar may emerge as a useful tool for analyzing citations in policy documents (Bakkalbasi, Bauer et al., 2006; Noruzi 2005). Neither measure is currently workable as an indicator, partly due to limitations in method and partly because policy documents do not always cite any underpinning evidence. This indicator could be applied at the individual level, but would be more useful at the group level and above since the number of citations for an individual is likely to be small. A benefit of analyzing citations in policy documents is that it could provide information on research from any of the four pillars that is influencing policy.

4.3.d. Health Impacts

It is very difficult to link research to health impacts (including attribution and time-lags to research impacts): it can require studies to link health and research indicators, rather than specific indicators for research impact on health. As such, our approach to the health impacts is to identify the most appropriate indicators for health that can potentially show a link to health research. Health impacts fall into three major groupings: health status, determinants of health, and health system performance (Statistics Canada and Canadian Institute for Health Information 2008). CIHI and Statistics Canada have already collected large volumes of data in each of these groupings.

Health Status Indicators

Improving the health status of individuals is not only the primary goal of health research, it is also the most likely to capture the imagination of public and policy maker alike. Measures of improved health

[18] There is a risk that the various media do not use the highest quality research available, since their role is to find "newsworthy" stories, rather than to inform the public about clinical research excellence. It would be interesting to cross-reference the coverage analysis with quality metrics to identify what proportion of the best research is covered in the media and what level of quality (for example, which percentile) major media stories cite.

are commonplace and form a vital part of understanding where to invest health care funding, and where to invest health research funding (through either directed funding or through the interest of researchers in specific health problems). There are essentially three ways to improve health: through reducing death (mortality), through reducing disease (morbidity), and by improving the quality of life of individuals (quality-adjusted mortality). As with all health impact indicators, the most appropriate health status indicators depend heavily on the aim of the research conducted. As such the "appropriate indicators" shown are classes, each with an example of the sort of information that would ideally be collected to establish the impact of the research on health.

MORBIDITY

Prevalence: Prevalence is the number of cases for a condition in a population (shown as a percentage of that population). Measuring prevalence allows us to track of the condition in question. Changes in prevalence can be related to research for that condition using retrospective studies. This indicator works for the population level, from subgroups to full population, and can use Canadian data that are already collected on a number of conditions (by, for example, the Public Health Agency of Canada – PHAC). As a condition-specific, rather than research specific-indicator, prevalence can be applied to research in any of the four pillars.

Incidence: Incidence is the number of new cases for a condition per 100,000 population. Incidence provides a different measure to the prevalence of a condition, since it provides an understanding of the rate of growth of that condition. As with prevalence, linking to research findings can be undertaken using retrospective studies. Again, as with prevalence, data are already collected on incidence for certain conditions (Statistics Canada 2008a). As an indicator, incidence is most useful at the population level (from subgroups to full population) and can be applied to research from any of the four pillars.

MORTALITY

Potential Years Life Lost (PYLL): PYLL represents the number of years of life lost due to premature death (before 75), and provides a measure of mortality than can be standardized across conditions. Data on PYLL are already collected (Statistics Canada 2007) and are most useful at the population level (from subgroups to full population). PYLL is applicable to research in all four pillars, as long as studies can provide a link between the changes in PYLL and the research findings.

QUALITY-ADJUSTED MORTALITY

Quality-adjusted Life Years (QALYs): QALYs provide a value between 1 (perfect health) and 0 (death) to represent quality of life for each year lived after a health intervention. While this is useful for building in a quality of life measure for medical interventions, it also makes them difficult to aggregate to general conditions. By virtue of their link to interventions, QALYs can more readily be linked to research, since research findings can be linked to interventions more easily than health impacts. The QALY approach can apply to any of the four pillars, although current QALY data are not as strongly linked to pillar IV research. QALYs are also collected in many different countries, making international comparisons of changes in quality-adjusted mortality possible for Canadians. In Canada, at a population level, QALY data are collected through the Canadian Community Health Survey.

Patient-reported Outcome Measures (PROMs): QALYs provide a "medical" view of improvements to health, using changes to the health outcomes for individuals. PROMs provide patient views, based on a standardized questionnaire about quality of care and quality of life post-treatment. PROMs are used in

the U.K. by the NHS as part of the outcome measures for improving health care (Fitzpatrick, Bowling, et al. 2006). Since this indicator uses the views of individual patients on their experience of outcomes, linking to research findings is currently a problem. Relating PROMs to research needs to be done at relatively high-levels (such as hospitals or disease states) that could be studied further to identify any links to research impacts. Although not currently in use in Canada, PROMs can provide a valuable alternative to the top-down approach to identifying health outcomes. PROMs are applicable to all pillars of research, as they do not relate to research directly, but rather to conditions and experiences of health care.

Determinants of Health Indicators

Determinants can be split into three major subcategories: modifiable risk factors (such as personal behaviour), modifiable social and cultural determinants, and environmental determinants (including the natural environment and the built or work environment). Since determinants of health can be very wide ranging, identifying specific indicators for every possible determinant is not possible in this report. Instead we identify examples of indicators for the different types of determinant.

Modifiable Risk Factors

Examples – obesity and alcohol consumption: Both of these are good examples of modifiable risk factors. Obesity can be modified in a number of ways and relates to a specific measure (BMI), while alcohol consumption is a good example of a single behaviour that can be modified. The measures of these modifiable risk factors can be useful for individuals, but are more useful at population levels. To link the risk factor indicators to research, the indicators need to link to the specific health problem under investigation in the research. The research that affects many of these modifiable risk factors is in pillar IV, but there are also be research findings from the other pillars that will link to changes in risk factor levels.

Social and Cultural Determinants

Examples - education levels and social cohesion: Indicators for social and cultural determinants must be specific to the determinant in question (for example, literacy levels for education), since trying to link the changes in a specific determinant to any health research can be very difficult. Not having a specific outcome measure would make it even more difficult. Potential areas where health research could influence social and cultural determinants include HSR research (suggesting the need for community health workers as an aspect of the health system), or through improving conditions for community health workers to work more effectively (Harter and Leier n.d.). The indicators for social and cultural determinants need to be compared by region, as aggregation to national levels could lose valuable context. It is possible that any health research could affect these determinants, but it is most likely that pillar IV will inform changes to large-scale social and cultural changes.

Environmental Determinants

Example – air pollution levels: Measuring levels of pollutant in air (parts per million) is a good example of an environmental determinant that can be affected by health research (for example, research on the effect of air pollution on respiratory diseases (D'Amato, Liccardi, et al. 2000)). Environment Canada maintains data on the levels of air pollution (Environment Canada 2005), and Statistics Canada has data on other factors such as second-hand smoke levels in Canada (Statistics Canada and Canadian Institute for Health Information 2008). Changes in the risk factor should be investigated for contributing factors in order to identify which research (if any) was a contributor to the change. As with other large scale determinants, air pollution levels should be considered by region, since

aggregation could lose information. Environmental determinants in the broadest sense (including changes to home and work environments) can be affected by research from all pillars, which would need to be taken into account when trying to link research to changes in the levels of any environmental determinant.

Health Care System Indicators

In Canada, changes to health system performance are typically considered to come under eight different factors: acceptability, accessibility, appropriateness, competence, continuity, effectiveness, efficiency, and safety (Statistics Canada and Canadian Institute for Health Information 2008). The data currently collected in these domains by Statistics Canada and CIHI are designed to provide information on some aspects of the health system, but are not comprehensive. Some of the indicators identified below attempt to broaden the health system indicators to cover all aspects of health care.

ACCEPTABILITY

Example – self-reported patient satisfaction: Acceptability is generally considered to be best measured by patient satisfaction with the health system (Canadian Institute for Health Information 1999). Surveying patients to identify their experience of the health service links in with the PROM measure used in health outcomes. Data could be taken directly from those questionnaires to determine the acceptability of the service provided to an individual (Niagara Health System n.d.). Some existing self-report surveys are not rigorous data collection tools and should be used with caution. Data from patient satisfaction surveys can be applied from health care provider level to regional, but would not be very useful beyond regional levels since information would be lost in aggregation. As the aim of HSR is to improve health service delivery, this indicator is primarily relevant to pillar III research.

ACCESSIBILITY

Example – wait times: Wait times for specific conditions and/or interventions are already collected by CIHI (Canadian Institute for Health Information 2008b), so this indicator would not require additional data collection. The issue with wait times is that they only apply to secondary care, and so neglect other aspects of the health system. Wait time data can be useful at provider, region, or population levels, but are not useful for individuals. Linking wait time data to research requires specific projects to understand the factors affecting wait times. This could link to any pillar of research, but is most likely to link to pillar III.

Example – appointment statistics: Statistics for time to appointments for different groupings (for example, socio-economic, gender, and ethnicity) could help to identify accessibility to primary care, which is something wait times cannot provide information on, but would require collecting data from primary care providers about 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 (Jones, Elwyn, et al. 2003). As with wait times, this indicator is most useful at provider, region, or population levels, and is most likely to link to pillar III research. To understand equity of access (an important part of having a fair health system), it is important to note that an analysis identifying gender, socio-economic groups, ethnicity, etc., for the groups accessing the health service (primary and secondary care) would need to be done.

APPROPRIATENESS

Example – adherence to clinical guidelines: Identifying the appropriateness of treatment is equivalent to asking, "Are health practitioners using the most appropriate methods and treatments in their work?" This can be done through clinical audit for health practitioners (Godwin 2001). Clinical audits can be used for individuals, but to evaluate changes in practice, it is most useful at provider, regional, or national levels. As a measure of the adherence to clinical guidelines, it is most likely that pillar III research will inform appropriateness, since the research informing the guidelines has already been captured through the "analysis of clinical guidelines" indicator in the informing decision making category.

COMPETENCE

Example – civil law suits against the health system: Competence is considered to be the appropriate application of skills in the health system (Canadian Institute for Health Information 1999), and can be approximated by identifying those times when the system is not appropriate, for example through counts of civil law suits against the health system. Counts of civil law suits by clinical area over time can be used to show data from individuals upward, depending upon the defendant involved in the suit. By monitoring the areas in which law-suits occur over time, it is possible to identify which clinical areas improve their performance. Linking these performance changes to research is difficult, however, and would require studies to identify the reasons for changes in levels of competence. Pillar III research is most likely to change competence, since HSR is the area of research most closely associated with service delivery.

CONTINUITY

Self-reported continuity of care: Continuity of practice is an important part of the health service, and can be 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" (Centre for Health Services and Policy Research 2004). Surveying patients to identify their perception of the continuity of their care provides a method for identifying experience for patient groups (Centre for Health Services and Policy Research 2004). An alternative to surveying patients is to use administrative data. In this case, the data collection is simple, but the results often do not identify well with experience of continuity, which is the aspect of continuity that is desirable (Centre for Health Services and Policy Research 2004). Continuity of care data could be applied for individuals, health care providers, or regions, but are heavily reliant on self-reported data being collected in the same way across Canada. Continuity is most likely to be impacted upon by pillar III research, but could conceivably be affected by pillar II research as well.

EFFECTIVENESS

Example – re-admission rates: The numbers of re-admissions by condition over a set time period, year on year, can provide an indication of whether the care provided by the health system is effective (Canadian Institute for Health Information 2008a). The main issue with this indicator is that it can only provide information on conditions that require secondary care; there is currently no indicator that can provide a similar measure of effectiveness for primary care or social care. Re-admission rates are useful at provider, region, or population levels, but not for individuals. They are most likely to be linked to pillar III research, but could relate to other pillars for specific interventions.

EFFICIENCY

Actual versus expected hospital stay: Measuring the length of stay for a patient as compared to the expected stay for the condition can provide information about the efficiency of the secondary care provided, but it cannot take into account individual complications or co-morbidity. This indicator should only be used for provider, region, or national comparisons; it is not useful for individuals. This indicator is likely to be most useful for pillar III research, since it relates to the service provision in secondary care.

Cost input versus output: Collecting data on the inputs to health care services and on the different factors identified as outputs (for example, available beds, emergency admissions, etc.) provides information that can be fed into a stochastic model to identify efficiency, such as a Stochastic Frontier Analysis (SFA) (McGlynn, Shekelle, et al. 2008: Lordan 2007). A large volume of these data, such as the costs of health care provision and the different outputs of health care, are already collected for health care providers. The benefit of using input and output data and a model in which it can be assessed is that all aspects of the health system can be addressed (not just secondary care), since data on inputs and outputs can be collected for primary and social care services. These data should only be used at the level of health care providers. The link to research findings would need to be established for any changes in inputs/outputs through a separate analysis. It is most likely that pillar III would be the one to affect efficiency of services, since that is one of the aims of HSR.

SAFETY

Safety is defined as reducing the risks of an intervention or health care environment (Canadian Institute for Health Information 2008a; Canadian Institute for Health Information 1999).

Example – adverse drug effects: The numbers of adverse drug effects, with their year- on-year change, is an easily measurable safety issue, and one of the most visible to the public. By linking the changes in the numbers of adverse drug effects to changes in practice, it is possible to link to any research that may have led to a reduction (or otherwise) in effects. This indicator could also apply to adverse surgical effects or reactions to anaesthesia. This indicator applies to providers and at provincial and federal levels by aggregating provider data. Changes in adverse drug effects will most likely be due to research in pillars II and III.

Example – hospital-acquired infections (HAIs): By measuring the levels of HAI and monitoring the year-on-year change, it is possible to understand what effect is achieved by procedures designed (through extensive health research) to reduce HAIs. HAIs are a very current safety issue, and are easy to measure and link to specific policies and research findings. This indicator can be used at the provider, provincial, and federal levels. It is most likely to link to pillar III research on HAI reduction procedures, but could link to any pillar when the links between research and HAI reduction are investigated.

4.3.e. Broad Economic and Social Impacts

Economic and social impacts can be very broad. Economic impacts must examine more than simple commercialization of research, although this is an important economic impact). They need to account

for the economic impacts of performing health research (rather than funding other activities),[19] and for the "health benefit" of research (that the positive health impacts of research relate to specific costs of implementing research findings in the broad health system). There are also social impacts of health research, with changes to the well-being of the population and changes to society as a whole through improved population health (particularly where research for populations with specific health challenges is concerned, such as the aboriginal population of Canada).

Activity Impact Indicators

Performing health research can have benefits, not just through research findings, but through the activity of research itself. 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, the employment of and retention of clinical staff, the funding brought into the country or province from international firms, and the ability to use new knowledge from other studies (covered in "absorptive capacity").

Economic rent (labour rents): Economic rent is the economic benefit (in $) of employing people in health research rather than in another capacity. Economic rent can account for the differential impacts of human resources versus other industries, that is, the benefit of employing people in biomedical research as opposed to employing them in another field. This differential measure is the concept of labour rent, or the excess earnings over and above the marginal cost of the labour. For example if a researcher is paid $30,000 and the next best job he could get would pay $25,000, then the economic rent is $5,000 per annum (Garau and Sussex 2007). As an indicator, economic rent can be applied as long as baseline data on research inputs can be identified through Statistics Canada data (Science, Innovation and Electronic Information Division 2008), and the outputs data required could be captured through Rx&D (Rx&D 2007; Rx&D 2006). Calculating economic rent has been done for the pharmaceutical sector (Garau, Sussex 2007), and in the U.K. has recently been applied to research from any funding source (Buxton, Hanney, et al. 2008). It has proven to be a powerful method for identifying the economic benefit of the activity of research, but requires good-quality data on inputs and outputs and a level of analytical expertise to perform studies. Because economic rent involves a number of assumptions and data collection issues, it is sensitive to poor data if there are small numbers involved. As a result, this approach is really only applicable from the institution or organizational level upward, and would be strongly recommended at a national level. Based on the recent U.K. study, economic rent may also be applicable at the funder or disease area level (Buxton, Hanney, et al. 2008). Previous research has not split the economic rent on health research by pillars of health research, and to do so would require additional testing of the method.

Commercialization Indicators

Previous evaluations have generally used a bottom-up approach to understanding commercialization, using measures such as patent licences, research product sales revenues, and spin-out companies (Muir, Arthur, et al. 2005). This provides collectable data, but misses the other aspects of

[19] MacDonald and Knoppers (Appendix E) also stress this need to understand whether funding was used appropriately to fund health research rather than other endeavours, although they address it from an "ethical use of funding" standpoint rather than an economic one.

commercialization that occur because of research funding, for example, through publicly funded researcher consultancy to industry.[20] To try to understand the less measurable impacts of research on commercialization, one must look at the overall behaviour of the health products industry and its relationship with R&D through producer rents and spillover effects (Garau and Sussex 2007). Collecting information on both of these aspects of commercialization can provide an element of triangulation of evaluation findings.

Licensing returns: 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 (Science-Metrix 2008; Byrd 2002). This measure is not recommended for individuals, since the numbers of licensed patents are likely to be small for individuals, but is recommended for groups, institutions, provinces, and nationally. Licensing returns can be relevant to all pillars of research, but there is likely to be an emphasis on pillars I and II.

Product sales revenues: Sales revenues of products developed in Canada provide a simple measure of the economic impact of health products (Science-Metrix 2008, Byrd 2002). There are difficulties, however, in linking sales revenues to research findings due to the other factors that affect sales. This indicator is recommended for provinces and nationally, and could be used for specific funders. It is not recommended for individuals, groups, or institutions, since attributing sales to these levels would be very difficult. All pillars of health research could contribute to health products, but it is likely that there would be an emphasis on pillars I and II.

Valuation of spin-out companies: Using the portfolio values 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 is relatively easy to identify (Lonmo 2008) as is their valuation if they are publicly traded. Accurately valuing privately held biotechnology spin-outs is not usually possible, because the valuations are based on perceived value to a small set of financiers. Any use of these indicators over a period of years would have to take into account market conditions that could drastically change the value of public companies, and the perceived value of privately held companies; economic techniques can deal with these changes to conditions. This indicator is recommended for provinces and nationally, and could be used for specific funders. It is not recommended for individuals, groups, or institutions, since the valuations of companies are not useful in assessing the impacts of individuals or small groups. This indicator can be applied to all pillars, but would be likely have an emphasis on pillars I and II, where most spin-out companies occur.

Economic rent (producer rent and spillover effects): Producer rent is the economic benefit to a company over and above expected revenues. Spillover effects are the external effects of investing in R&D on groups that are not invested in (for example, investment from abroad in private R&D having benefits in Canada) (Garau and Sussex 2007). By collecting data on producer rents (through revenue statements and expected levels of profit) and identifying spillover effects, this indicator could identify

[20] To some extent, the presence of consultancy and interactions with industry is covered by an indicator on "numbers of consultancies to industry," but merely identifying the consultancy link does not provide us with a way to identify the economic impact of that consultancy.

the broader impacts of research on commercial outcomes (including the impacts of publications, networking, increasing absorptive capacity, and entrepreneurship). Spillovers can be identified through an analysis similar to that used in the recent analysis of the economic impact of medical research in the U.K. In this analysis, the researchers identified public and private investments in R&D[21] 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 (Buxton, Hanney, et al. 2008). Calculating producer rent and spillovers requires an understanding of the economic techniques underpinning analysis, making it a likely organizational, institutional, or national indicator, not suitable for smaller funders or organizations. Producer rent and spillover effects can provide information for provinces and nationally, and could potentially be used for specific funders, but could not be used for the impacts of individuals, groups, or institutions. This indicator can be used for any pillar of health research, but it is likely that the research informing producer rents and spillovers will be from pillars I and II, as these are the pillars most closely linked to the health products industry.

Health Benefit Indicators

Identifying health gain was discussed in the health benefits section. Previous evaluations have used various measures of health gain to estimate the economic benefit of improved health and quality of life (Buxton, Hanney, et al. 2004). Most recent studies in this area have focused on total health benefits, since this provides a more equitable view of economic benefits than using those through the improved production function of healthier workers (Buxton, Hanney, et al. 2008). It is important to understand not only the benefits of improved health, but also how they relate to the costs of improving health. We suggest using the net benefit of health as an approach to understanding health gain benefits. In essence, this 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). This measure can be used for both indicators of quality-adjusted mortality (broad scope health improvement) identified in the health impact indicators.

Health benefit in QALYs per health care dollar: Using QALYs gained, divided by the cost of achieving that gain, to identify the net improvement in health is a useful approach that has been used elsewhere (Buxton, Hanney, et al. 2008), because QALYs can be monetized (using a controversial methodology, for example, (Bingham 2001)). This approach provides a monetary net benefit that can be compared to other uses of capital (other than on health research). Using QALY values presents three main problems for identifying accurate economic returns. First, the economic value of a QALY, although building on a body of research on valuing health, still relies on a debatable economic value. Second, QALYs are only calculated for some treatments, so this economic approach is not comprehensive. Third, the use of an analysis of QALY valuations requires expert research, and would have to be outsourced to a group that is capable of performing such an analysis. As an overview measure of the changes to large-scale health and one that requires extensive input and output data on health research costs and improvements in health, this indicator is not recommended for individuals or groups, but can be very useful for institutions, funders, provinces, and nationally. This approach could also be used for any of the research pillars.

[21] Complementarities have been shown to exist between public and private funding for R&D (Congressional Budget Office 2006)

Health benefit in PROMs per health care dollar: As with the use of QALYs, 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, PROMS have not been used in this kind of approach before (so there is no data to compare to), and second, PROMs have not been monetized, so this measure can only be compared to other PROM measures. As with the QALY per $ measure, this indicator is not recommended for individuals or groups but is useful for institutions, funders, provinces, and nationally. It can also be applied to all pillars.

Well-being Indicators

Well-being is a difficult concept to measure. Here, we try to use a comprehensive measure and two aspects of well-being to highlight the importance of measuring well-being as an impact of health research.

Annual report of Human Resources and Social Development Canada (HSRDC): HRSDC has multiple indicators that can be used to identify well-being, however, they currently have no links to research (health or otherwise) except through the "health" section of the well-being indicators (covered in the "health impacts" category). Since these data are already collected and publicly accessible, it is an attractive starting point to access information about changes in well-being in Canada. Because of the difficulty in attributing changes to research findings, this indicator is only recommended at the national level. It is possible that any health research could affect well-being (indeed likely), but pillar IV research is probably most easily linked to changes in well-being at a population level.

Happiness: There are a number of measures used to assess happiness. We recommend using established survey techniques for happiness/depression, such as the self-report happiness scales used by Statistics Canada and the short depression/happiness scale (Joseph, Linley, et al. 2004). 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. This indicator is recommended for provinces and nationally, but not for assessment of research by individuals, groups, or institutions. It can apply to research in any pillar, since we currently have no way to effectively link this impact to health research.

Level of social isolation: By using loneliness scales to measure the social isolation of individuals (such as the UCLA loneliness scale (McWhirter 1990)), we can access information about the social isolation of individuals. As with happiness, linking changes in social isolation to health research is currently very difficult, so this measure should only be used to track levels of social isolation provincially and nationally. It is not recommended as an indicator for the research of individuals, groups, or institutions. As with happiness measures, this indicator could link to any pillar of research.

Social Benefit Indicators

It is important to remember that the final outcomes of research can be social impacts as well as economic ones. Social benefits arising from health research can be incredibly wide ranging, so we provide an example of an indicator, rather than indicators for all possible social impacts.

Example – socio-economic status: The causality of socio-economic status to health outcomes is well known (Ostry 1999; Kelley and Hurst 2006), but it is not understood whether health research can alter socio-economic status. Data that identifies the socio-economic status of individuals in Canada should be collected in order to identify whether changes in socio-economic status correlate with research

impacts, which is something that would have to be assessed through specific research studies. This information is useful for provinces and nationally, but is not be recommended for addressing the research impacts of individuals, groups, or institutions. Since the link of research to changes in socio-economic status is difficult to show, it is possible that research from all pillars could impact, but the most likely research area to directly impact on socio-economic status is pillar IV.

4.3.f. Theoretical examples of indicator sets for evaluation

Using only single indicators provides a skewed view of the impacts of research: it is vital that impacts be assessed through *sets* of indicators. Here we create three theoretical examples of how to create indicator sets to illustrate the uses of a set of indicators to answer specific evaluation questions for different funding bodies: an academic biotechnology cluster; a federal funder's fellowship funding; and a provincial funder's project grant funding.

Example 1: Biotechnology

Biotechnology has key aims in producing high-quality academic research, translating research to industry, and producing returns on the research industry uses. Therefore, the primary evaluation questions for the biotechnology cluster are likely to be:[22]

- Have we produced the best research?
- Have we successfully translated our research to commercial entities?
- Have we facilitated a commercial gain through our research findings?
- Have we created employment opportunities for our graduates?

For each of these questions, there are sets of indicators from our "appropriate indicators" that can help to provide complete answers.

For the first question, some advancing knowledge indicators can be used to showcase quality while the relative citation impact and the number of highly cited publications can provide information on the international competitiveness of the research. It would also be useful to find out where else that research is occurring in Canada, so using the share of publications indicator could help identify which other groups in Canada are producing similar research. Although this indicator only identifies where publications are being produced, modifying it to address only the share of HCPs across Canada (with the data that are available through the HCP analysis) will show where research is being produced and what proportion of high-quality Canadian research is produced by the cluster.

To answer whether there has been an influence on industry, information about the interactions with industry can be measured through the use of a variety of indicators: first, through co-authorship analysis to identify where research links with industry are; second, through levels of additional funding brought in from industry sources; third, through the number of licensed patents held by the cluster; and fourth, through the consultation to industry that is performed by researchers in the cluster.

[22] This does not mean to imply that there would not be other questions of interest, such as whether new researchers had been trained and what they are doing, whether research had been effectively translated to the public, and whether research results had affected future research directions for funders.

Measuring the commercial gain from the research, question three, uses indicators on the commercialization of research results (broad economic and social benefits). Measuring the funds from licensing of patents and valuing any spin-out companies arising from the cluster could provide indicators of value on the outputs of the research. To complement those measures, a study to identify the producer rent and spillovers from commercial research associated with findings from the cluster would provide a measure of return that could take into account how the research has affected the company performing the research. It would be unwise to simply sum these values to show a commercial return value, since the interactions between these factors would not be known.

Example 2: Federal fellowship funding

Funding individuals rather than research topics (be they programs or projects) often have an explicit aim to improve the research capacity for Canada and to improve opportunities for individual researchers. The primary evaluation questions likely for a federal fellowship scheme are:

- Have we increased the skill set of Canadian health research?
- Have we increased the number of skilled researchers working in Canada?
- Are our fellows producing high quality research?
- Are our fellows disseminating their findings to a variety of appropriate stakeholders?

To answer the first of these questions we need to be able to assess the researchers individually, identifying the areas of research they are working in, any research degrees they (or indeed members of their team) have been awarded, and any new awards or positions of prestige they have attained (these could be research awards, journal editorship, or keynote speaker activities). To identify areas of research, bibliometric analysis of research outputs by individual can provide an idea of the areas they work in. Research degrees awarded can be tracked through the capacity building indicator investigating research degrees. Prestige awards can be obtained through analysis of a common CV and/or an end-of-fellowship report.

Increasing the number of skilled researchers working in Canada can be achieved through analyzing the Statistics Canada data on researchers in health related subjects in a year on year fashion (capacity building). Attributing any change to the specific fellowship funding will be difficult, but it is highly likely that there has been a contribution.

Identifying the quality of the research outputs of fellows can be done using the advancing knowledge bilbliometric indicators – relative citation impact (adjusted for field, self-citations, numbers of publications etc.) and numbers or proportions of the worlds highly cited publications. Of course, these quality measures cannot be claimed exclusively by the fellowship funding, since a proportion of the research activity undertaken by researchers will be funded by project/program grants.

Disseminating to appropriate stakeholder groups can be identified through the informing decision making indicators; specifically through an analysis of media mentions, consultancies to industry/policy makers/advocacy groups etc., and through public lectures given. As with prestige indicators, these are most easily collected through a common CV or end-of-fellowship reporting survey. The specific desired stakeholder interactions will depend upon the nature of the fellowship and the research undertaken.

Example 3: Provincial project grant funding

Grant funding differs from fellowship funding in a number of its goals, as does provincial from federal funding. These are reflected in the evaluation questions that would be most likely for a provincial grant funding stream:

- Have we produced high quality research?
- Have we improved the research endeavour in our province?
- Have we encouraged investment in our province?
- Do we manage to translate our findings to improve health in the province?

Identifying high quality research can be achieved through the advancing knowledge indicators of relative citation impact (adjusted for field, self-citations, numbers of publications etc.) and numbers (or proportions) of the worlds highly cited publications (HCPs). For a provincial funder, the comparator for proportions of HCPs would be against other provinces, a comparison that can be made based on the analysis of author addresses. It would also be prudent for any province investigating the quality of their research to normalize the findings by any of a number of factors such as research dollars spent, numbers of funded researchers etc. This allows comparisons across provinces that vary in the size of their research endeavour.

Improving research across the province can be addressed through improved capacity for research. The specific indicators to collect here would be around improvements in the personnel capacity in the province, improvements in the receptor capacity of the province and any improvements in the infrastructure for research in the province (using infrastructure funding as a proxy indicator for this improvement).

Provincial funding agencies can also require information on how their investment has improved investment in the province. Investment can be: a research investment from federal funders (such as CIHR) which is a capacity building indicator; an investment in provincial research findings through patent licensing (economic benefit); or a direct investment by pharmaceutical or medical device companies or venture capitalists in the province (as measured through clustering of companies in the province and through spin-out valuation in the province).

Measuring the translation of research findings to improved health is very difficult, but indicators of improved health for specific conditions related to provincial research can be used to assess improvements that could be linked to research (measured through improvements in QALYs, PROMs and decreasing prevalence and incidence of conditions). Improvements in well-being should also be considered here using the recommended indicators of happiness and social isolation.

Overview of examples

Clearly these examples do not cover the full spectrum of potential evaluations, but they do highlight the importance of addressing specific evaluation questions. Without first tailoring the evaluation questions to the research/funding under investigation, the evaluation itself will be aimless and is unlikely to be cost-effective in identifying the impacts of the research. Questions need to be tailored to the aggregate level of the research, the aims and mission of the funding, and the research subject areas under investigation.

5. Chapter 5: Conclusions

Our mandate from sponsors for this work was to define and understand "return on investment" in the broadest sense, and to provide a menu of options for data collection that can be used by anyone to create a rigorous, robust, and practical evaluation for his or her organization. These options needed to be evidence based and to incorporate both quantitative and qualitative approaches to data gathering. To build such a menu requires an understanding of the impacts (the "returns") that health research can have and an understanding of the pathways to those impacts.

Currently, there is no evaluation framework that adequately captures the impacts and pathways to impacts that health research across Canada can have. By investigating the wealth of evaluation frameworks that exist (Appendix C, p. A232) the panel has identified (a) what knowledge can be applied in a Canadian context, and (b) what is missing, in order to create an evaluation framework for Canadian health research (Figure 9).

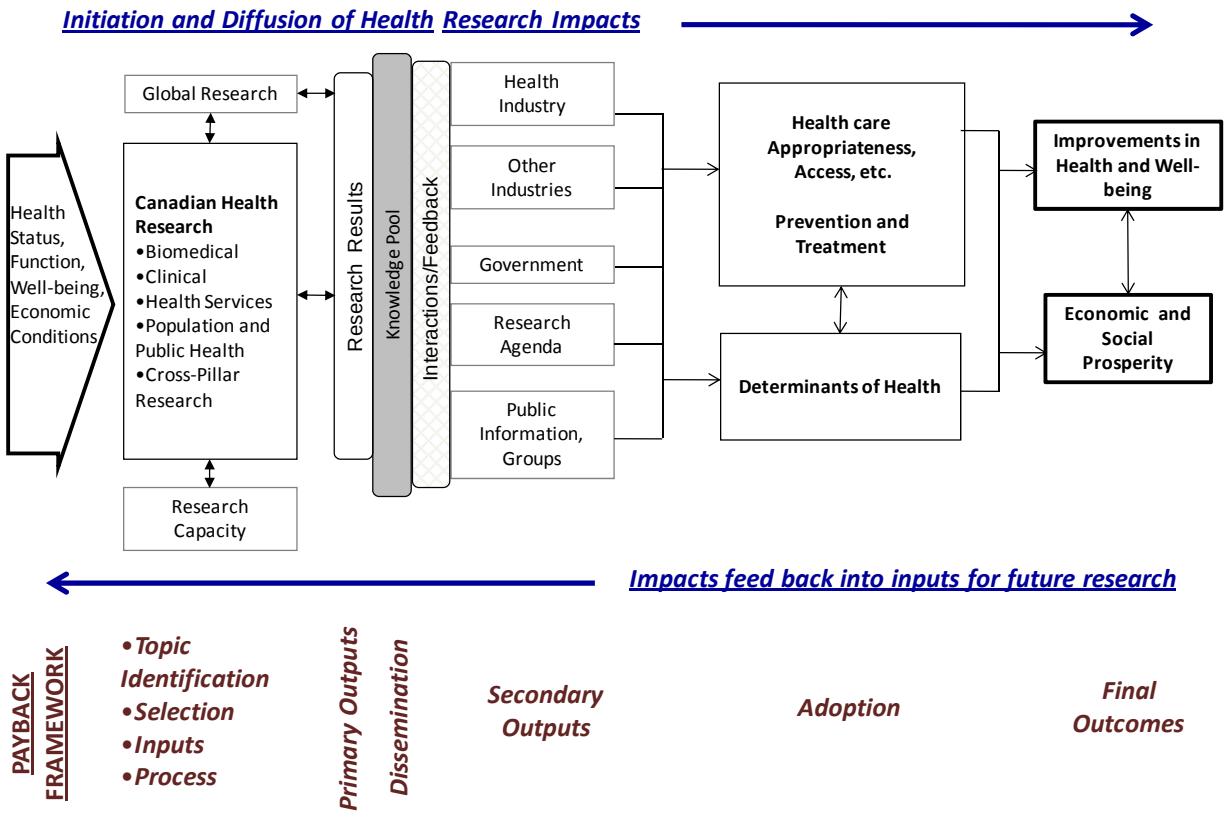


Figure 9. A simplified version of the CAHS evaluation framework developed by the panel to understand the processes and pathways to health research impacts.

This framework provides the basis for identifying:

- where data should be collected
- what data should be collected
- how to perform evaluations of health research to ensure comparability and comprehensiveness

What it does *not* do is provide an instruction manual for research funders to follow verbatim. Depending upon the reason for evaluating health research (be it accountability, advocacy, or learning for the funder), there are different sets of evaluation questions that need to be considered. Evaluations of research funding need to take into account:

- the mission of the funding organization or funding stream
- the needs of key stakeholders in the evaluation
- the costs of data collection for the evaluation

This means that every evaluation is slightly different. By using an underlying framework to show the logic of the evaluation and comparable indicators of impact, however, evaluations can be customized for the purposes of both the funding organization and the Canadian health research system as a whole.

Having a research evaluation framework does not provide a "magic bullet" that suddenly allows all evaluations to perfectly identify all impacts arising from health research, or how much impact can be attributed to specific projects. Rather, the framework provides a starting point from which to understand how the research undertaken has gone on to have the impacts—including understanding the attribution or contribution of the research to a particular impact, the likely difference that the research has made (when compared to a counterfactual), and an understanding of the time-lags involved in research impacts.

Any evaluation that uses the framework can only be as good as the data it collects and the assumptions that surround the use of those data. For example, if the data collected on health changes related to the research funding only includes information about the health of a proportion of individuals, then the evaluation will not show the true impacts of that research funding, and this reduces the potency of the evaluation. The data must also be accurate and precise if the evaluation is to be credible. If the assumptions surrounding the data are incorrect (for example, if the health change measured is not something that can be affected by the research under evaluation), then the evaluation can be dismissed as not relevant.

It is practically impossible to identify the most appropriate data to collect for any potential evaluation of health research in Canada. The panel has instead identified a "menu" of potential indicators that can be brought together into sets for specific evaluations (shown in the table in the executive summary on pages 25-32). These indicators cover five impact categories (advancing knowledge, capacity building, informing decision making, health impacts, and broad economic and social impacts) and conform to criteria on the attractiveness and feasibility of any single indicator (Butler 2008). When used together in sets, indicators can help to create focused, appropriate, balanced, robust, integrated,

and cost-effective evaluations (HM Treasury, Cabinet Office, et al. 2001). If evaluators of Canadian health research use these "appropriate indicators," then Canada can begin to create a large collection of comparable data for use in benchmarking the performance of different stakeholders in the health research process (for example, researchers and funders).

A number of recommendations have arisen (see the "Recommendations" chapter) from the process of producing this assessment. These recommendations cover:

- what framework and indicators should be used in Canada
- how to use the evaluation framework to improve the identification of health research impacts and pathways to impacts
- why international efforts are needed to improve indicators, understanding of attribution, learning from evaluation findings, and standardizing data collection
- what Canada can do to continually improve on "research on research" to track attribution, trace returns, and understand and improve the system of knowledge translation
- adopting the selected framework and indicators in Canada to allow comparisons of impacts across different funders (nationally and internationally)

Implementing these recommendations will be an ongoing process in Canada, requiring a combined effort from all health research stakeholders. Research funders need to implement appropriate data collection techniques, researchers need to commit to maintaining data on their broad outputs and outcomes (that is, taking responsibility for the broader impacts of their research, not just the publication of results), public stakeholders (such as patient groups) need to be incentivized to provide information on the research they use in decision making, and researchers into evaluation need to commit to improving tools for evaluation.

As evaluations become more sophisticated, other indicators will need to be developed, validated, and added to the library of indicators that we have assembled as of 2008. Many other indicators can be imagined, but improving the evaluation tools for addressing health research impacts is not something that can be achieved without concerted effort and dedicated funding. By creating a research discipline that is dedicated to understanding how research goes on to have impacts (research on research), Canada can influence international work on evaluating research impacts. By better understanding the process of evaluation and how to attribute impacts to research, we can begin to answer questions about the links between outcomes and research inputs, and also about the cost-effectiveness of evaluation itself.

There also needs to be collaboration in Canada on the maintenance of data for all research funders. For example, one of the recommended indicators is an analysis of media mentions for research. This indicator could potentially benefit all funders of research, but would create duplication of effort if all funders were to maintain records of media mentions. Identifying a group to take responsibility for collecting and housing this data then becomes an important part of making this evaluation system work for Canada.

If Canada can implement these recommendations and begin to collect data on research impacts, the benefits can be substantial: improving the cost effectiveness of the health research system, improving

the health outcomes arising from research, and leading the way internationally in “research on research”. At a time of economic uncertainty, knowing how to fund effectively and efficiently can give Canada a significant edge in health R&D.

6. References

Genome vikings. Editorial. 1998. *Nature Genetics,* **20**, pp. 99-101.

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.

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).

BAMBERGER, M., 2006. *Conducting Quality Impact Evaluations Under Budget, Time and Data Constraints.* Washington, D.C.: World Bank.

BARBER, H.D. and CRELINSTEN, J., 2004. *The Economic Contribution of Canada's R&D Intensive Enterprises 1994-2001.* Toronto, Ontario: Research Infosource Inc.

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.

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.

BIRDSELL, J.M. and MATTHIAS, S., 2001. *S.E.A.R.C.H. Program Evaluation, Research and Development Blueprint.* Alberta, Canada: On Management Ltd.

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.

BUTLER, L., 2008. Metric selection criteria: ongoing research. Personal Communication.

BUTLER, L., 2002. A list of published papers is no measure of value. *Nature,* **419**(6910), pp. 877.

BUXTON, M. and HANNEY, S., 2005. Development of a multidimensional 'payback' framework for assessing investments in health research. *Italian Journal of Public Health,* **2**, pp. 60.

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. 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.

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 ACADEMY OF HEALTH SCIENCES, 2007. *Prospectus for a Major Assessment: The Return on Investments in Health Research: Defining the Best Metrics.* Ottawa, Ontario: CAHS.

CANADIAN BROADCASTING CORPORATION, January 4th, 2007 - last update, doctors weigh potential U.S. stem cell research funding changes [Homepage of CBC], [Online]. Available: Doctors weigh potential U.S. stem cell research funding changes [August 25th, 2008].

CANADIAN HEALTH SERVICES RESEARCH FOUNDATION, 2007. *Canadian Health Services Research Foundation (CHSRF): 2007-2009 Strategic Plan.* Ottawa, Ontario: CHSRF.

CANADIAN HEALTH SERVICES RESEARCH FOUNDATION, 2005. *Annual Report.* Ottawa, Ontario: CHSRF.

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, 2007. *National Health Expenditure Trends: 1975-2007.* Ottawa, Ontario: CIHI.

CANADIAN INSTITUTE FOR HEALTH INFORMATION, 1999. *National Consensus Conference on Population Health Indicators*. Ottawa, Ontario: CIHI.

CANADIAN INSTITUTES OF HEALTH RESEARCH, May 14th, 2008a - last update, clinical research initiative (CRI): A partnership for clinical research [Homepage of CIHR], [Online]. Available: http://www.cihr-irsc.gc.ca/e/22113.html [August 18th, 2008].

CANADIAN INSTITUTES OF HEALTH RESEARCH, 2008b. Payback categories. Ottawa, Canada: CIHR.

CANADIAN INSTITUTES OF HEALTH RESEARCH, 2007a - last update, 2007-2008 CIHR grants and awards guide. Available: http://www.cihr-irsc.gc.ca/e/22630.html#1-A1.1.

CANADIAN INSTITUTES OF HEALTH RESEARCH, November 2007, 2007b - 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 OF HEALTH RESEARCH, September 4th, 2007c - last update, open access to health research publications: CIHR unveils new policy [Homepage of CIHR], [Online]. Available: http://www.cihr-irsc.gc.ca/e/34851.html [August 26th, 2008].

CANADIAN INSTITUTES OF HEALTH RESEARCH, September 17th 2007, 2007d - last update, an overview of CIHR [Homepage of Canadian Institutes of Health Research], [Online]. Available: http://www.cihr-irsc.gc.ca/e/30240.html#slide1_e [June 26th, 2008].

CANADIAN INSTITUTES OF HEALTH RESEARCH, August 14th, 2007e - last update, research funding overview [Homepage of CIHR], [Online]. Available: http://www.cihr-irsc.gc.ca/e/22955.html#expenditures [August 25th, 2008].

CANADIAN INSTITUTES OF HEALTH RESEARCH, 2005a. *Canadian Institutes of Health Research: Performance Report for the Fiscal Year 2004:2005.* Ottawa, Ontario: CIHR.

CANADIAN INSTITUTES OF HEALTH RESEARCH, 2005b. *Developing a CIHR Framework to Measure the Impact of Health Research.* Ottawa, Ontario: CIHR.

CANADIAN LONGITUDINAL STUDY ON AGEING, 2007 - last update, Canadian longitudinal study on ageing: overview [Homepage of Canadian Longitudinal Study on Ageing], [Online]. Available: http://www.clsa-elcv.ca/en/overview/ [October 18th, 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.

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.

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.

DAHLGREN, G. and WHITEHEAD, M., 1991. *Policies and strategies to promote social equity in health: Background document to WHO – Strategy paper for Europe.* Stockholm, Sweden: Institute for Futures Studies.

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.

DEVELOPMENT ASSISTANCE COMMITTEE WORKING PARTY ON AID EVALUATION, 2002. *Glossary of Key Terms in Evaluation and Results Based Management.* Paris, France: OECD.

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.

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.

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.

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].

EVANS, R.G. and STODDART, G.L., 1990. Producing health, Consuming health care. *Social Science & Medicine,* **31**(12), pp. 1347-1363.

FAST, E., 2007. *Mobilizing Science and Technology: The New Federal Strategy.* PRB 07-32E. Ottawa, Ontario: Library of Parliament.

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.

FUNDING FIRST, 2000. *Exceptional Returns: The economic value of America's investment in medical research.* New York, New York: The Lasker Foundation.

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.

GIBBONS, M. and GEORGHIOU, L., 1987. *Evaluation of Research. A Selection of Current Practices.* Paris, France: OECD.

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].

GOVERNMENT OF CANADA, 2007a. *Mobilizing Science and Technology to Canada's Advantage.* Federal Government of Canada.

GOVERNMENT OF CANADA, 2007b. Mobilizing Science and Technology to Canada's Advantage: Summary.

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.

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.

HANNEY, S., HOME, P.D., FRAME, I., GRANT, J., GREEN, P., and BUXTON, M., 2006. Identifying the impact of diabetes research. *Diabetic Medicine,* **23**(2), pp. 176-184.

HANNEY, S., MUGFORD, M., GRANT, J., and BUXTON, M., 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**, pp. 937-947.

HARTER, J. and LEIER, M., n.d. *No Time To Care: Community Health Workers and the New Economy.* Vancouver, B.C.: Simon Fraser University.

HEART AND STROKE FOUNDATION CANADA, 2008 - last update, about us [Homepage of HSFC], [Online]. Available: http://www.heartandstroke.com/site/c.ikIQLcMWJtE/b.3479075/k.C8B0/About_Us.htm [June 26th, 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 24th, 2008].

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.

JOINT ECONOMIC COMMITTEE, 2000. *The benefits of medical research and the role of the NIH.* Washington, D.C.: US Senate.

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.

KAPLAN, R.S. and NORTON, D.P., 1992. The Balanced Scorecard - Measures that Drive Performance. *Harvard Business Review,* (January-February), pp. 71-79.

KELLEY, E. and HURST, J., 2006. *Health Care Quality Indicators Project Conceptual Framework Paper.* OECD Health Working Papers No. 23. Paris, France: OECD.

KHOO, K., COLUCCI, R., HRYNIUK, W., KAMINO, R., REDINA, T., and SAVAGE, C., 2007. The Cost of Cancer Drugs in Canada. *Report Card on Cancer in Canada,* **10**, pp. 53-59.

KING, D., 2008. *Doctoral Graduates in Canada: Findings from the Survey of Earned Doctorates, 2004/2005.* No. 065. Ottawa, Ontario: Statistics Canada.

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).

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.

LAURIER, C., March 14th, 2005 - last update, program evaluation and corporate governance at social development canada - A hard look at how we made it work: speech by the deputy minister, SDC to the executive forum on evaluation. Available: http://www.hrsdc.gc.ca/en/cs/comm/sd/deputy_minister/speeches/2005/050308.shtml [October 27th, 2008].

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.

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.

MAREDIA, M., BYERLEE, D., and ANDERSON, J., 2000. *Ex Post Evaluation of Economic Impacts of Agricultural Research Programs: A Tour of Good Practice.* Rome, Italy: Consultative Group on International Agricultural Research.

MARTIN, B.R., 1996. The use of multiple indicators in the assessment of basic research. *Scientometrics,* **36**(3), pp. 343-362.

MCCLELLAN, M. and HEIDENRICH, P., 2003. Biomedical research and then some: the causes of technological change for heart disease. In: K.M. MURPHEY and R.H. TOPEL, eds, *Measuring the gains from Economic Research: An economic approach.* Chicago, Illinois: University of Chicago Press, pp. 163-205.

MCDAID, D. and COOKSON, R., 2003. Evaluating health care interventions in the European Union. *Health Policy,* **63**, pp. 133-139.

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.

MICHAEL SMITH FOUNDATION FOR HEALTH RESEARCH, December 6th 2007, 2007 - last update, about us [Homepage of MSFRH], [Online]. Available: http://www.msfhr.org/sub-about.htm [June 26th, 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., 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.

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.

MURPHY, K.M. and TOPEL, R.H., September, 1999-last update, the economic value of medical research [Homepage of University of Chicago], [Online]. Available: http://faculty.chicagogsb.edu/kevin.murphy/research/murphy&topel.pdf [September 30th, 2008].

NASON, E., 2008. *Health and Medical Research in Canada.* 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 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 INSTITUTE OF CHILD HEALTH AND HUMAN DEVELOPMENT, September 16th, 2008 - last update, SIDS: "back to sleep" campaign [Homepage of National Institutes of Health], [Online]. Available: http://www.nichd.nih.gov/sids/ [October 14th, 2008].

NATIONAL INSTITUTES OF HEALTH, 2008. *Revised Policy on Enhancing Public Access to Archived Publications Resulting from NIH-Funded Research.* NOT-OD-08-033. Bethesda, Maryland: NIH.

NIAGARA HEALTH SYSTEM, n.d. - last update, quality 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.

O'NEILL, M.F., 2007. Biotechnology in Ireland: hard work, money and the brain gain. *Drug Discovery Today*, **12**(3-4), pp. 108-111.

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, 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, 2007. *OECD Science, Technology and Industry Scoreboard 2007: Innovation and Performance in the Global Economy.* Paris, France: 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.

OSAMA, A., 2006. *Multi-Attribute Strategy and Performance Architectures in R&D: The Case of The Balanced Scorecard*, RAND Corp. PRGS.

OSTRY, A., 1999. *Explaining the Link Between Socioeconomic Status and Health.* HPRU 99:12D. Vancouver, B.C.: Centre for Health Services and Policy Research.

PEACH, I. and MARSHALL, J., 2008. *The Economic Benefits of Health Research in Saskatchewan.* Regina, Saskatchewan: Saskatchewan Institute of Public Policy.

PICARD, A., 2008. Five cents of every research dollar goes to cancer research – is it enough? *The Globe and Mail,* **August 26th**, pp. L4.

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.

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.

REHN, C., KRONMAN, U., and WADSKOG, D., 2007. *Bibliometric indicators: definitions and usage at Karolinska Institutet.* Version 1. Stockholm, Sweden: Karolinska Institutet University Library.

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.

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, January 23rd, 2006 - last update, did you know? our community [Homepage of Rx&D], [Online]. Available: http://www.canadapharma.org/DidYouKnow/index_e.html [July 27th, 2008].

SASKATCHEWAN HEALTH RESEARCH FOUNDATION, January 2007. *Evaluating Saskatchewan's Health Research Strategy: An Evaluation Framework.* SHOULD THIS BE LISTED BELOW THE NEXT ENTRY -- 2008 BEFORE 2007?

SASKATCHEWAN HEALTH RESEARCH FOUNDATION, 2008 - last update, about SHRF: overview [Homepage of SHRF], [Online]. Available: http://www.shrf.ca/Default.aspx?DN=9cc6cc2c-ff73-4b42-ab8d-1e02d36da1fe [26th June, 2008].

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].

SCIENCEWATCH.COM, n.d. 2008 - last update, country profiles 2008: canada. Available: http://sciencewatch.com/dr/cou/2008/08junCanada/ [August 12th, 2008].

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].

SMUTYLO, T., 2001. *Crouching Impact, Hidden Attribution: Overcoming Threats to Learning in Development Programs.* Ottawa, Ontario: Evaluation Unit, International Development Research Centre.

SOLAR, O. and IRWIN, A., 2006. Social determinants, political contexts and civil society action: a historical perspective on the Commission on Social Determinants of Health. *Health Promotion Journal of Australia*, **17**(3), pp. 180-184.

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, April 28th, 1999 - last update, national graduates surveys - public use microdata file. Available: http://www.statcan.ca/bsolc/english/bsolc?catno=81M0011X [November 5th, 2008].

STATISTICS CANADA and CANADIAN INSTITUTE FOR HEALTH INFORMATION, 2008. *Health Indicators 2008.* Ottawa, Ontario: CIHI.

TERRY, R., 2005. Funding the Way to Open Access. *PLoS Biology*, **3**(3), pp. e97.

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].

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.

UBC FACULTY OF MEDICINE, August 12th, 2008 - last update, media coverage [Homepage of University of British Columbia], [Online]. Available: http://www.med.ubc.ca/news/media_coverage.htm [August 21st, 2008].

UK CLINICAL RESEARCH COLLABORATION, 2007. *Health Research Classification System.* London, England: UKCRC.

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. 2008 - 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. , 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 DRIEL, R., April, 2002 - last update, stem cell research: A double edged sword [Homepage of Association for Reformed Political Action], [Online]. Available: http://arpacanada.ca/index.php/issuesresearch/41-pro-life/82-stem-cell-research-a-double-edged-sword [August 25th, 2008].

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.

WAHL, A., 2008. The power of innovation. *Canadian Business,* **March 17th**.

WEISS, A.P., 2007. Measuring the Impact of Medical Research: Moving From Outputs to Outcomes. *American Journal of Psychiatry*, **164**(2), pp. 206-214.

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.

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.

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.

WOODWARD, G., MANUEL, D., and GOEL, V., 2004. *Developing a Balanced Scorecard for Public Health.* Toronto, Ontario: ICES.

ZAHRA, S.A., and GEORGE, G., 2002. Absorptive Capacity: A Review, Reconceptualization and Extension. *Academy of Management Review,* **27**(2), pp. 185-203.

7. Biographies for Panel Members and Staff

Assessment Panel Members

Cyril Frank (Chair), University of Calgary

Cy Frank is an Alberta Heritage Foundation for Medical Research Scientist; the McCaig Professor of Joint Injury and Arthritis Research and Executive Director of the Alberta Bone and Joint Health Institute; and a Professor in the Division of Orthopaedics at the University of Calgary. He was the inaugural Scientific Director of the Institute of Musculoskeletal Health and Arthritis of the Canadian Institutes of Health Research from 2000 to 2006. Over the years, Dr. Frank has chaired or been a member of many local, national, and international committees for many different associations, including the Arthritis Society, Canadian Institutes of Health Research (CIHR), and the Canadian Orthopaedic Association. He has currently co-authored over 225 peer-reviewed publications, 39 book chapters, and presented 364 abstracts locally, nationally, and internationally. His research has focused on ligament repair, ligament transplantation, and on the mechanisms of osteoarthritis. He is currently the leader of a multicentre AHFMR Team grant entitled: "Creating Bone and Joint Health from the Bedside to the Bench and Back Again – Designer Therapies to Reduce the Burden of Osteoarthritis (OA) – from Mechanisms to Prevention."

Dr. Frank was a member of the Canadian Institutes of Health Research's "Key Results" Working Group from 2004 to 2006 and a member of CIHR's Standing Committee on Performance Measurement, Evaluation, and Analysis from 2002 to 2006. He served as an external reviewer of the RAND report, "The Returns of Arthritis Research," prepared for the U.K. Arthritis Research Campaign in 2004. In addition, he was one of the architects of the model and pilot project known as "Alberta Bone and Joint Health Institute Hip and Knee Pilot Project," which was presented to the House of Commons in 2006.

Renaldo Battista, Université de Montréal

Dr. Renaldo Battista (MD, MPH, ScD, FRCP (C)) is Professor and Director of the Department of Health Administration (DASUM) at the Université de Montréal. From 1982 to 2003, he was on staff in the Departments of Epidemiology and Biostatistics, and of Medicine at McGill University. He was president of the Quebec Council for Health Technology Assessment (CETS) from 1994 to 2000, and President and CEO of the Quebec Agency for Health Services and Technology Assessment (AETMIS) from 2000 to 2004. His most significant research contributions can be grouped under three headings: integration of preventive services with clinical practice; development and implementation of clinical practice guidelines; and health technology assessment. Together, they testify to a continuous and evolving interest in

knowledge transfer or the creation of effective linkages between the production of scientific information and its use by health professionals, managers, and decision makers.

Dr. Battista was a member of the Advisory Board of the Institute of Health Services and Policy Research of the Canadian Institutes of Health Research from 2001 to 2003. He has been a member of the Scientific Advisory Board of Health Canada since January 2005, and has held the Canada Research Chair in Health Technology Assessment (HTA) since October 2005. He was elected a member of the Canadian Academy of Health Sciences in 2006.

Linda Butler, Australian National University

For the last seven years, Linda Butler has been a research Fellow and head of the Research Evaluation and Policy Project (REPP) at the Australian National University. She is a leading authority on bibliometric analysis, and has been engaged in the theoretical and empirical study of research evaluation for nearly 20 years. Her findings on the effects of using publication counts to determine the distribution of research funding to universities have had significant public policy impact.

With the impending changes to the way research funding is distributed to universities in Australia, Ms. Butler's research on quantitative indicators has become of critical importance. She was appointed to chair the Department of Education Science and Training's working group on quality metrics in 2006, and has subsequently been contracted by the Australian Research Council to continue advising on quality metrics through to 2009 as the new assessment process, "Excellence in Research for Australia," is developed and implemented. From 2004 to 2007, Ms Butler was a member of the Australian National Health and Medical Research Council's (NHMRC) working group on Measures of Research Impact and Achievement (MORIA). This group was established by the NHMRC's Research Committee to devise measures of research impact and achievement that can be applied across its funding portfolio, covering commercialization and health gains, in addition to knowledge creation.

Her research interests include assessing quantitative measures of research performance, with a particular interest in measures for the humanities, arts, and social sciences; using bibliometric techniques to map the organizational structure of Australia's research landscape, and to analyze its relative strengths and weaknesses; and the assessment of the impact of research in the broader community outside of academia.

Martin Buxton, Brunel University, United Kingdom

Martin Buxton is a Professor of Health Economics at Brunel University. He is also the Director of the Health Economics Research Group, which he established and that is now one of the leading university-based units for health economic evaluation research in the U.K. He has over 30 years' experience in the methods and practice of economic evaluation in health care, and has been involved in a large number of studies, including: international drug trials, evaluations of national programs for heart and liver transplantation, screening for breast cancer and abdominal aortic aneurysms, evaluation of pilot projects such as the picture archiving and communication system (PACS) at Hammersmith Hospital, assessments of the impact of medical audit, and other organizational changes. He has been a member of the Appraisal Committee of the National Institute for Clinical Excellence, and has advised most of the major pharmaceutical companies. A consortium led by Dr. Buxton has just completed a quantitative study for the Medical Research Council, Wellcome Trust, and the Academy of Medical Sciences, which estimates the economic benefits from medical research in the U.K. in the fields of cardiovascular disease and mental health.

Dr. Buxton's research interests include: the refinement of quantitative and qualitative methodologies for economic evaluation of all types of all health technologies through empirical application; practical policy uses of evaluation; and assessment of payback from health services research.

Neena Chappell, University of Victoria

Neena L. Chappell, PhD, FRSC, is the Canada Research Chair in Social Gerontology, Professor of Sociology and Centre on Aging, at the University of Victoria. She was founding Director of the Centre on Aging at the University of Manitoba (1982–1992), and the first Director of the Centre on Aging at the University of Victoria (1992–2002), developing both into world-class research facilities, while ensuring accessibility to the community. For more than 25 years, she has been a leader in gerontological research, focusing on three areas: quality of life for seniors, care-giving, and the health care system and related policy. Her research can be characterized as partnerships with other researchers in the social sciences, and with non-researchers in government and community agencies.

Dr. Chappell promotes relevance and scientific rigour. She conducts large, quantitative studies; small, in-depth qualitative studies; and publishes policy papers and commentaries on the current societal situation. She argues for recognition of the positive aspects of aging, as well as recognition of problems and challenges.

She has written more than 250 academic articles and reports, authored books, and spoken extensively on health, health care policy, and formal and informal care-giving that has relevance to seniors. She is an editorial board member of the Journal of Aging and Ethnicity, Canadian Journal of Sociology, Journal of Aging Studies, Journal of Applied Gerontology, Social Sciences, Journal of Gerontology, and The Gerontologist, and has sat on many boards and committees. She was an executive member of the Canadian Institutes of Health Research Task Force and its interim governing council. She is currently a member of Health Canada's Science Advisory Board and is a fellow of the Royal Society of Canada.

Sally C. Davies, Department of Health, United Kingdom

Professor Sally Davies is the Director General (DG) of Research and Development for the Department of Health (DH) and National Health Service (NHS). As DG, she developed the new government research strategy, "Best Research for Best Health," with a budget of £1 billion, and is now responsible for implementing the National Institute of Health Research (NIHR). She is a board member of the Office for the Strategic Coordination of Health Research (OSCHR). Together, the DH and NIHR fund research (applied, clinical and policy), research infrastructure, research capacity development, and support the NHS in its research activities. Key national programs include the internationally recognized "Health Technology Assessment Programme, Service Delivery, and Organization," and "Invention for Innovation." New initiatives include the recently announced Biomedical Research Centres and Units, and Experimental Medicine and Cancer Research facilities. Sally led the U.K. delegation to the WHO Ministerial Summit in November 2004, spoke on R&D at the World Health Assembly in May 2005, is a member of the WHO Global Advisory Committee on Health Research (ACHR), and is currently chairing the Expert Advisory Committee for the development of the WHO research strategy.

Aled Edwards, University of Toronto

Aled Edwards, PhD, is a noted Canadian structural biologist. He is Banbury Professor in the Banting and Best Department of Medical Research at the University of Toronto, and Director and CEO of the international Structural Genomics Consortium (SGC). The SGC is an Anglo-Canadian-Swedish, public-private partnership created to increase the number of protein structures relevant to human health that are available in the public domain. In each of the past two years, the SGC has deposited over 20% of the new human protein structures to the public database.

Dr. Edwards co-founded Affinium Pharmaceuticals, a Toronto-based anti-infectives company; Jaguar Biosciences, a diagnostic company focused on hepatitis C; and Scate Consultants Inc., a company that commercializes bioremediation intellectual property. He also serves as a scientific advisor on the Canadian television drama "ReGenesis." He has served in management and advisory capacities for several biotechnology companies and international research consortia.

Chris Henshall, University of York, United Kingdom

Dr. Chris Henshall, MA, PhD, is Pro-Vice-Chancellor for External Relations at the University of York, where he works to promote innovation on a local, national, and international level. He graduated from Cambridge and Nottingham universities with a background in natural sciences, child development, and developmental psychology. Previously, he worked as the Director of the Science and Engineering Base Group in the United Kingdom's Office of Science and Technology (OST), which has a mandate to provide policy-makers and the public with independent analysis of public policy issues related to science and technology.

He is an experienced educator and researcher, having lectured part-time at the University of Maryland European Division, the University of Warwick, and the University of Southampton. Dr. Henshall has had a longstanding involvement in work that assesses the returns on investment from research. He commissioned research and convened an international seminar on payback from health research during his time at the Department of Health, and has subsequently served on steering committees for further government and charity studies and initiatives to assess the returns from health research and from research in general. Some of Dr. Henshall's past positions include: Deputy Director of Research and Development in the Department of Health (United Kingdom), Program Director for the Cambridge-based Health Promotion Research Trust, Principal Scientific Officer for the U.K. Medical Research Council, and Assistant Secretary in the Research and Development Division of the Department of Health.

Yann Joly, Université de Montréal

M. Yann Joly, LLB, LLM, is a distinguished lawyer and project manager at the Centre de recherche en droit public at the Université de Montréal. He is currently an ethical and legal consultant in the biotechnology industry, providing advice on socio-ethical issues related to biotechnological research. He has collaborated with industry and currently sits on the Ethics Advisory Board for Genizon BioSciences Inc., where he advises on ethical practices and policies.

M. Joly is the North American coordinator of the Association for Research and Formation in Medical Law. He worked as a sessional lecturer at McGill University, where he taught intellectual and industrial property in the Faculty of Law. M. Joly sat as a legal advisor on several McGill University Health Centre ethics committees from 2004 to 2006. He was a member of the External Working Group on the Registration and Disclosure of Clinical Trial Information (EWG-CT), created by Health Canada in 2006. M. Joly is currently a member of the Regenerative Medicine Economic, Legal, and Social Issues Network (RMEthnet); the Association for Research in Medical Law (ARFDM); and the Quebec Network of Applied Genetic Medicine.

Gretchen Jordan, Sandia National Laboratories, U.S. Department of Energy

Gretchen Jordan, BA, MA, and PhD, is a principal member of the technical staff in the Science and Technology Strategic Management Unit at Sandia National Laboratories, a U.S. Department of Energy (DOE) laboratory. She has worked with both basic and applied research programs at Sandia and DOE headquarters since 1993. Dr. Jordan's expertise is in assessing research program effectiveness, performance measurement, and finding innovative ways to improve the efficacy of scientific research and technology programs. She is an expert in logic models and program theory, with a focus on developing evaluation frameworks and strategy.

A ground-breaking leader in R&D evaluation, Dr. Jordan has published in the field, and chairs the American Evaluation Association's Topical Interest Group on Research, Technology, and Development Evaluation. She is on the steering committee of the Washington Research Evaluation Network (WREN). In 2007, she was elected a Fellow of the American Association for the Advancement of Science (AAAS), for measuring performance and diversity within organizational structures and developing innovative methods for assessing the effectiveness of research programs.

Dr. Jordan's background is in mathematics and economics and she collaborates with colleagues at the Center for Innovation at the University of Maryland. Prior to joining Sandia National Laboratories, she held both academic and government positions. She chaired the Business Administration Department at the College of Santa Fe and was a staff member of U.S. Senator Pete V. Domenici, as well as serving on the U.S. Senate Budget Committee.

Terence Kealey, University of Buckingham, United Kingdom

Terence Kealey graduated in medicine from St Bartholomew's Hospital Medical School, London University, in 1975. He then specialized in biomedical research, achieving his DPhil (PhD) from the Department of Clinical Biochemistry, Oxford University, in 1982.

His biomedical research has focused on the cell biology of human skin. He has also studied the economics of science and higher education. His 1996 book, *The Economic Laws of Scientific Research*, argues that there is no market failure in science, and that it can be safely entrusted to the free market. His latest book, *Sex, Science, and Profits*, was published in 2008.

Dr. Kealey was appointed Vice-Chancellor the University of Buckingham in 2001. He writes occasionally for *The Times* and other British national broadsheets.

Michael C. Wolfson, Statistics Canada

Dr. Michael C. Wolfson, BSc (Toronto – mathematics, computer science, and economics), PhD (Cambridge – economics), is Assistant Chief Statistician, Analysis and Development, at Statistics Canada.

Dr. Wolfson concentrates on health statistics and analytical and modeling programs. His areas of expertise include program review and evaluation, tax/transfer policy, pension policy, income distribution, design of health information systems, microsimulation modeling of socio-economic policy and health, and analysis of the determinants of health.

He has held positions in the Treasury Board Secretariat, the Department of Finance, the Privy Council Office, the House of Commons, and the Deputy Prime Minister's Office prior to joining Statistics Canada. He was also a Fellow of the Canadian Institute for Advanced Research Program in Population Health (1988–2003). His numerous articles have addressed topics such as assessing the inter-generational equity of Canada's pension and health care systems, the design of an appropriate system of health statistics, modeling disease determinants and treatments, inequality trends in Canada and the United States, polarization and the decline of the middle class, measurement of economic growth, and income and income inequality as determinants of population health.

Dr. Wolfson has been active in research councils and institutes such as the Canadian Institutes of Health Research (CIHR), the Research Council of the Canadian Institute for Advanced Research (CIAR), the WHO Advisory Committee on Health Monitoring and Statistics, and the Council of the Canadian Population Health Initiative of the Canadian Institute for Health Information (CIHI).

He has participated in peer reviews for the National Institutes of Health (NIH), Canadian Foundation for Innovation, and presented at the 2004 and 2007 Organization for Economic Co-operation and Development (OECD) World Forum on Measuring Progress, and was appointed to the Federal Public Service Pension Plan Advisory Committee in 2007.

Steven H. Woolf, Virginia Commonwealth University, United States

Steven H. Woolf, MD, MPH, is Professor in the Departments of Family Medicine, Epidemiology, and Community Health at Virginia Commonwealth University. He received his MD in 1984 from Emory University and did his residency in family medicine at Virginia Commonwealth University. Dr. Woolf is also a clinical epidemiologist, with training in preventive medicine and public health from Johns Hopkins University, where he received his MPH in 1987. He is board certified in family medicine, in preventive medicine, and in public health.

Throughout his career, Dr. Woolf has worked to promote effective health care services and to advocate the importance of health promotion and disease prevention. In recent years, his work has turned to social determinants of health. Dr. Woolf has conducted studies that show that addressing poverty, education, and the causes of racial and ethnic disparities could accomplish far more to improve the health of Americans than investing predominately in medical technological advances. In addition to publishing more than 100 articles, he has tried to bring this message to policymakers and to the public through testimony in Congress, editorials in major newspapers, and speeches.

Since 1987, Dr. Woolf has served as science advisor, member, and now senior advisor to the U.S. Preventive Services Task Force. Dr. Woolf edited the first two editions of the *Guide to Clinical Preventive Services*, and is the author of *Health Promotion and Disease Prevention in Clinical Practice*. He is associate editor of the American Journal of Preventive Medicine, and served as North American editor of the British Medical Journal. He has consulted widely on matters of health policy with government agencies and professional organizations in the United States and Europe, and in 2001 was elected to the Institute of Medicine.

ROI Assessment Staff

Rhonda Kennedee – Meeting and Events Coordinator

Rhonda Kennedee has been planning and staging events and meetings for 25 years. Her experience includes large and small meetings, as well as major conferences for the Calgary Health Region and the Division of Rheumatology. Rhonda has also worked in research at the University of Calgary, as a study coordinator for a number of rheumatological studies, including the Silicone Breast Implant study in 1993, and a study to determine whether exercise can make a difference in controlling fibromyalgia pain. As well as planning events and meetings, Rhonda has worked as a news producer and writer for several Canadian television networks.

Linda Marchuk – Research Associate

Linda Marchuk (BSc, RT) has over 35 years' experience as a research technologist and manager. She currently manages several research projects for Dr. Cy Frank in the Faculty of Medicine at the University of Calgary, where her primary interests are in the facilitation of musculoskeletal research in the field of osteoarthritis. Prior to joining Dr. Frank in Calgary, she worked as a transplantation immunology research technologist in the Faculty of Medicine and the University of Alberta Hospital in Edmonton. She has authored 44 manuscripts, 3 book chapters, and 81 scientific abstracts.

Edward Nason – Health Research Evaluation Analyst and Writer

Prior to joining this CAHS assessment, Edward was a policy analyst at RAND Europe. His main area of research has been health R&D policy, and his experience spans the full timeline for research, from planning R&D spending through to evaluating outcomes. He has worked for numerous international funders, including payback evaluations for the Canadian Institutes for Health Research, the Irish Health Research Board (HRB), the Heart and Stroke Foundation of Australia, and the U.K. Economic and Social Research Council. He has also advised the British Department of Health on how to use bibliometric techniques to facilitate short listing for R&D funding decisions.

Larissa Sommerfeld – Research Assistant

Larissa holds BAs in both International Relations and Political Science from the University of Calgary. She has participated in study programs in Cuba, Quebec, and the Czech Republic. Larissa has had the opportunity to work for New York Congresswoman Carolyn B. Maloney, and has also volunteered with local Alberta politicians. She joined the ROI team as a research assistant after working as a contract employee for the Faculty of Medicine at the University of Calgary.