Despite intense interest in defining the social, health and economic impacts of health research investments globally— and in Canada—as proof of value-for-money, no validated method for measuring return on investments yet exists. Until now, issues of complexity combined with major gaps in methodology have limited the ability to link health research products to outcomes at a relevant level (e.g., to be useful to stakeholders: individual funders, decision-makers, institutions, researchers or clinicians). In this article, we discuss current approaches to measuring returns on investment, analyze key issues and gaps that need to be bridged to improve returns on investment, and present a new method that may help overcome them.

Two approaches used to define return on investment for health research

As recently reviewed by Shiel and DeRuggerio for the Canadian Academy of Health Sciences, 2 main approaches have been used over the past 20 years to measure return on investments. The first uses what can be called “top-down” econometric calculation. The Lasker Foundation commissioned health economists to use this approach to estimate return on investments from health research in the United States. They monetized improvements in life expectancy and quality of life between 1970 and 1990 by ascribing to them a value of roughly $1.5 trillion/year. They then attributed roughly one-third of those gains (about $500 billion) to health research, and concluded that these economic returns far exceed the costs of the health research that contributed to them (> 20-fold).

Adaptations of this method were also used more recently to estimate the returns on investment of research in Australia in 2003 and 2008. Both reviews suggested that more than double any investment in research is being returned as a benefit to Australian society.

The second approach, a “bottom-up” approach, to measuring return on investments has been evolving for over 2 decades through the efforts of Buxton and Hanney and coworkers. Their “payback model,” recently adopted and modified by the Canadian Institutes of Health Research in Canada, has involved logic-model tracking of new knowledge in phases from knowledge production at the researcher level adding to a knowledge pool, and then from there to secondary outputs and adoption to final outcomes. They classified impacts into 5 categories: knowledge production; research targeting, capacity and absorption; informing policies and product development; health and health sector benefits; and broader economic benefits. These categories move “downstream” from research through translation to society and can capture outputs of interest for different audiences. For example, demonstrating research-related changes in the health and health sector benefits category would be of interest to clinicians. As one example of its use, Wooding and colleagues evaluated the payback profiles of various programs for the Arthritis Research Campaign in the United Kingdom using quantitative scales. They concluded that there is variability in outputs of research projects, short-term projects with flexible funding (e.g., investigator-initiated research projects) provide excellent value for money invested, individuals translate research findings, and, although their outputs
are also highly variable (some produce very little), funding some research institutes can potentially produce the greatest impact in all 5 domains per pound invested.22

A hybrid version of the payback approach that uses some econometric assumptions combined with more detailed data validation and linkage to research outputs has been used to estimate the recent economic benefits and health gains from investments in cardiovascular and mental research in the United Kingdom between 1975 and 1992.24 They have calculated ranges that estimate that for cardiovascular disease, 10%–25% of health care benefit is attributable to United Kingdom research, with an average time lag of 17 years. By monetizing these health gains and combining them with estimated gross domestic product gains of 30% that include “spillover” effects to other sectors, they cautiously estimate a total annual internal rate of return of £0.39 for every £1 of public or charitable money invested.

**Issues to resolve in determining return on investments from health research**

For both the econometric and payback methods of estimating return on investment in health research suffer from some common issues that need to be resolved in order for their results to be validated and for their results to be useful to each stakeholder audience. Table 1 shows 4 evaluation examples with some of their strengths and weaknesses.

The first major hurdle for both methods is how to deal with the so-called “attribution issue” that plagues all research.25 That is, the inability to determine the exact contributions of health research (versus other factors) in achieving its end goals: positive changes in health, health care or adding substantially to social and economic prosperity. The econ-

### Table 1: Approaches to the measurement of social, health and economic returns on investment in health research

| Country                        | Perspective                  | Methods                                                                 | Major findings                                                                 | Strengths                                                                 | Weaknesses                                                                 |
|-------------------------------|-----------------------------|-------------------------------------------------------------------------|--------------------------------------------------------------------------------|--------------------------------------------------------------------------|---------------------------------------------------------------------------|
| United States                 | National level              | Econometric modelling; investments and monetized health benefits pooled over 20 years | < $25 billion in investment contributed to about $500 billion in estimated health improvement | • Low cost to perform<br>• Valuable at policy level<br>• Easy to understand | • Top-down attribution assumptions<br>• Gross averages only<br>• Not useful at the funder level |
| Ireland                       | Program level               | Payback model applied to 8 diverse cases in areas of pain, obstetrics, heart attacks, dentistry and neurobiology | All 8 cases demonstrated some impacts and 6 cases produced impacts in 3 or more payback categories | • Detailed results; can track outputs and outcomes<br>• Can inform program improvement | • Small number of cases may not be generalizable<br>• Labour intensive; includes researchers<br>• Costly to evaluate per case |
| Australia                     | National level              | Econometric modelling with retrospective analysis over blocks of time. Monetized to DALYs and value of statistical life-year | Each dollar invested in Australian Health research and development returned $2.17 in health benefits, on average. Rate of return exceeded only by mining and retail | • Economic output is appealing<br>• Easy to understand results<br>• Good at the policy level<br>• Estimates time lags | • Top-down approach using temporal differences in mortality and morbidity<br>• Several assumptions can be criticized<br>• Not useful at the program or funder level |
| United Kingdom                | National level              | Estimates of research investments and monetized QALY benefits using a bottom-up approach (payback analogy) based on evidence from 46 different patient indication–treatment combinations, 1985–2005 | 39% internal rate of return on cardiovascular research investments. Return on mental health research was similar. Impacts lag by about 17 years. | • Sound basis for empirically estimating returns and time lags<br>• Economic outputs and outcomes are clear<br>• Includes “spillover” to research and development and other sectors, and counts contributions to GDP<br>• Based on bottom-up data | • Some calculations may be controversial (e.g., QALY as opportunity cost)<br>• Some data missing (resulting in extrapolations) and quality of some data uncertain<br>• Higher cost<br>• Historic data only, with no projections<br>• Not calculated at the individual funder level |

Note: DALY = disability-adjusted life-year, GDP = gross domestic product, QALY = quality-adjusted life-year.
metric method attempts to factor the other contributors out through assumptions. The payback approach can potentially isolate health research impacts more effectively depending on what metrics it uses and how specific its output assessments are, but it has also relied on some educated assumptions about contributions of research to the changes observed.

The second issue, which is related to the first, is known as the “counterfactual.” What would have happened if the research had not been conducted and how can this be determined? What controls for the lack of health research exist? Have the best controls been used for each research finding? Measuring impacts assumes that research has helped create them without necessarily linking them. That is clearly an assumption of the econometric method in that it had to ascribe an estimated and clearly debatable percentage of impact to research (33%). The payback method could suffer from the same issues, depending on what denominators and controls are selected, but, as it is measuring from the bottom up, it is likely better able to track impacts as they occur.

The third major issue is the well-known time lag for knowledge translation and how to account for it. From basic discovery to an effective therapy can take anywhere from 2 years to 30 years, adding to the difficulty of assessing returns at any single point in time. Even short-term longitudinal studies can thus also miss important impacts. The econometric method actually pools impacts across a large enough timeframe (e.g., over 25 years) that it presumably averages impact across enough time to capture a spectrum of individual research impacts at various points in their evolution. The payback method can use timescales, and it could use longitudinal, as opposed to cross-sectional, analysis so it could potentially determine time lags. The UK cardiovascular payback example, however, used a retrospective review approach with some assumptions and estimated an average time lag of 17 years, but the authors were quick to qualify that as an estimate only.

The fourth issue is the need for clear definition of what to measure and how to really capture impacts in meaningful terms. Economic measures are obviously numerical and easier to defend as metrics, whereas social and health improvements are more qualitative. The latter problem has been overcome to some extent by the use of measures that attempt to quantify improvements objectively (e.g., quality-adjusted life years) and potentially monetize them. There are many other, equally important nonnumerical “indicators of impact” that have required definition and validation, along with the metrics of impact, using criteria of attractiveness and feasibility.

Unique strengths and weaknesses of both approaches can easily be summarized. Gross econometric methods are clearly useful for advocacy and decisions at policy levels, and they are certainly more cost-effective, as they use only high-level administrative data to draw conclusions. They are less helpful at im-

Figure 1: Simplified schematic (top) of the health research impact framework developed by the Canadian Academy of Health Sciences, showing specific areas and target audiences where health research impacts can be found. The payback model, which identifies the stages of knowledge production from secondary outputs through adoption to final outcomes, is shown on the bottom. Reproduced with permission from the Canadian Academy of Health Sciences.
proving individual program planning or delivery at either institutional or funder-specific levels as they pool research impacts rather than teasing them apart. The payback method, on the other hand, while having the advantage of being a bottom-up approach, uses information from the researcher level downstream, which makes it less prone to assumptions about linkages to outputs and confounding effects, can suffer from being too detailed, slow and costly. It can certainly trace impacts across specific domains and time scales right down to an individual project level, but this is also much more labour intensive.

What would define the best method of measuring return on investments in health research?

To be practical and useful, the best method for measuring return on investment in health research must address all the issues noted above. It must also be useful to a full range of funders and research types, compatible with existing systems, suitable for international comparisons, and able to identify the full spectrum of potential impacts.26 The best method must also be feasible, not too labour intensive, and economically viable. It should be as accurate and responsive as possible within a reasonable evaluation budget that should represent a small percentage of the money invested in the research being assessed.

A new approach for addressing the gaps and needs

In early 2007, as part of its mission to provide unbiased analysis of important topics of international interest,27 the Canadian Academy of Health Sciences established a blue-ribbon panel of experts to create a simple but robust composite method that might address all of the challenges noted above.28 After considering the world literature and the options, including a review by RAND Europe29 of the strengths and weaknesses of many existing evaluation frameworks, they recommend a method that builds on the advantages of the “payback model”18–22 but adapts it to target-specific impacts in multiple domains at multiple levels (Figure 1). It is more of a bottom-up approach than the econometric approach, and, as in the payback model, it combines an impact category approach with logic model. This adds the advantage of specific program and project comparisons to impact framework models, such as the balanced scorecard, which can evaluate progress against targets for an organization but cannot explain how impacts occurred or how to improve them (Table 2).

| Model                        | Purpose                              | Major assumptions                                      | Uses                                         | Strengths                             | Weaknesses                                      | Validation          |
|------------------------------|--------------------------------------|-------------------------------------------------------|----------------------------------------------|----------------------------------------|------------------------------------------------|---------------------|
| Econometrics                 | To define return on investment at the macro level | • Data sources are good and complete <br> • Attribution is valid <br> • Values of improvements are valid | • Global assessment of return on investments <br> • Can be used for programs at high level | • Cost effective for countries <br> • Easy to understand <br> • Defines return in true economic terms | • High level <br> • Averages over time <br> • No help in improving outputs or outcomes | Funding First12      |
| Payback (logic-model and impact-category based) | To define impacts of projects (or groups of research projects, research programs or higher) in 5 categories | • Researcher feedback is valid <br> • Output data are valid | • Project and program evaluation and comparison of programs | • Useful for funders  <br> • Can define project returns in 5 categories at specific times or over time <br> • Excellent for comparisons <br> • Can inform improvement | • Complicated to perform <br> • Only as good as the data on the indicators and metrics <br> • Data intensive <br> • Expensive | Wooding et al.30 <br> Health Economics Research Group, et al.31 |
| Balanced scorecard (and other research-impact-based frameworks) | Accountability for outputs in a few domains <br> Target setting | • Impacts being measured are the important ones <br> • Data are valid and specific | • Defines several categories of impact and measures against those | • Focused on a few perspectives and a few outcomes for organizations <br> • Can assess progress against objectives <br> • Good in terms of accountability and advocacy | • Does not define how impacts occurred <br> • Does not necessarily help improve outputs or outcomes | Beaudet32 |

*None of the models have been rigorously validated, but all are in common use for various purposes, which suggests that each has value for some stakeholders.
The new model is specifically designed to track impacts in 5 categories: advancing knowledge, capacity building, informing decision-making, health benefits, and broad economic and social benefits (that can potentially include cultural outcomes). It was designed as a “roadmap of impacts version of the payback model” to help identify where proximal impacts can occur — the health industry, other industries, government, research decision-makers, or the public or public groups — and follows them distally through stages of adoption to final outcomes in health, well-being and social or economic prosperity. For the sponsors, the framework was also designed to trace research impacts in any or all impact categories for any of the 4 “pillars of health research” in Canada (basic biomedical, applied clinical, health services and systems, and population health) or at any level (from individual projects to the whole country).

**How to use the framework**

Several steps are required to use the new framework appropriately. The first is defining evaluation questions. The framework is then used to determine which categories of impact are expected and where they might occur. One example of how to use the framework and where to look for the various categories of impact is provided for clinical research (Figure 2). The next key step is the strategic selection of sets of impact indicators, chosen from a starting menu of 66 preferred indicators designed to an expert panel (e.g., areas where data are not yet being collected or where other barriers to measurement currently exist). The panel also provided a template for designing new indicators as a step toward building and expanding a library of validated health research impact indicators, with appropriate references. This will provide maximum flexibility and allow users to devise customized indicators for their specific needs while still meeting the required standards. This recommended-indicator resource should also help users progressively resolve complexities in assessing health research impacts.

**What the framework can and cannot do**

The framework can be used to do some things that were previously not possible. The 66 indicators in 5 domains of evaluation allow stakeholders to answer a wide variety of questions that address the value of research and research funding. Evaluation indicators (some of which are quantitative metrics) have all been validated and meet internationally accepted standards of attractiveness and feasibility. An example showing the evaluation of a research program aimed at developing high-quality research capacity for a provincial funder is shown in Table 3. This example uses the model to evaluate the effectiveness of research capacity development by 1 open and 2 different strategic research programs. These indicators could be used to compare capacity development per dollar but are not prescribed to do so.

Some aspirational indicators were also suggested by the expert panel (e.g., areas where data are not yet being collected or where other barriers to measurement currently exist). The panel also provided a template for designing new indicators as a step toward building and expanding a library of validated health research impact indicators, with appropriate references. This will provide maximum flexibility and allow users to devise customized indicators for their specific needs while still meeting the required standards. This recommended-indicator resource should also help users progressively resolve complexities in assessing health research impacts.

![Figure 2: Example of where the various impacts of clinical research can occur using the Canadian Academy of Health Sciences model.](image-url)
questions about the impact of research can range in scope from individual projects to the national level. The evaluation framework can also be adapted to various evaluation perspectives, from the individual to the international perspective. The framework can also be used as a roadmap to help identify specific impacts and, perhaps, even examine the impact of a specific piece of research, research program, body of work or funding entity as its relevant products are translated downstream.

The multidimensional nature of the framework should enhance the comprehensiveness of any evaluation undertaken as well as enable its users to capture the various strengths and weaknesses of the entity under evaluation.

Finally, the framework can describe feedback loops that may stimulate new research. It can assist in understanding and distinguishing the separate research effects from the influences of confounding factors, such as changes in the environment (e.g., a recession).

In general, tailored evaluations such as the modified payback framework enhance one’s ability to discriminate between 2 or more programs. However, the enhanced flexibility comes at a cost. The multitude of indicators may make it more difficult to synthesize results from different studies if very different indicators are used. Multiple indicators may also provide confusing results if some show improvement and others worsen. Resolving discordant results within the same evaluation or between evaluations can certainly be an important limitation.

Any evaluation will also be limited by the domains and indicators included in it. The 66 validated indicators currently listed do not cover the full spectrum of possibilities. New indicators will need to be developed and validated to further enhance the overall usefulness of the method and its comprehensiveness. Clearly, the payback framework cannot track impacts in domains outside the boundaries that it has established, nor can it identify impacts in audiences that have not been identified.

Challenges to be overcome

The designers of the framework recognize several important implementation challenges. First, funders will have to become familiar with the framework and indicators and how to apply them. This will take time and require dedicated people in various organizations to make an effort to implement the framework for their own purposes. Collaboration among organizations to achieve all of the practical prerequisites for any meaningful evaluation of impact on a national scale (including reliable and standardized data sources and data collection methods) will also pose some major challenges. Expanding the library of validated indicators to address all of the potential evaluation questions will be a challenge.

An even bigger challenge for all users of the new framework will be defining impact questions more explicitly, prioritizing them, and choosing sets of indicators and metrics that

| Table 3: Theoretical comparison of the impact on development of short-term research capacity of 3 programs at Agency A* and Agency B |
|---------------------------------------------------------------|
| Program                                         | No. of publications in high-quality outlets first-authored by a PhD student per grant per year | No. of PhD students who graduate per grant per year | No. of MD-PhD students who graduate per grant per year | Infrastructure and equipment grants (external funding attracted); $ per grant per year |
| Agency A: Individual operating grant program that can support trainees | 1.7 | 1.1 | 0.2 | 10 000 |
| Agency A: Strategic grant program for teams that can support trainees | 2.5 | 2.2 | 0.5 | 50 000 |
| Agency A: Technology commercialization grant program that can support trainees | 0.5 | 0.5 | 0 | 20 000 |
| Agency A: Average for the 3 programs | 1.6 | 1.2 | 0.2 | 17 000 |
| Agency B: Average for comparable programs | 3.5 | 2.7 | 1.2 | 150 000 |

*This evaluation shows that Agency B builds short-term research capacity best. To also measure research quality, the agency selected a metric that counts the number of publications of students in high-quality journals. These results suggest superiority of the strategic grant program, but Agency A lags behind Agency B in terms of these indicators. As this evaluation would be bottom up (data from researchers), the agencies could also identify which research projects and mentors were “best” at these outputs and possibly learn from them. By adding another indicator of publication quality (e.g., frequently cited publications) and 1 of quantity (publication counts), they could also determine which program advances knowledge to the greatest extent and express that in terms of dollar invested. Alternatively, they could look for longer-term impacts, such as number of trainees who get academic appointments (using an indicator to be developed). With investment data added, this information could be used to determine return on investment in capacity development as well as to help set program priorities in the future.
address the questions in a practical way. This sounds much easier than it is in practice, as experts note that there are almost an infinite number of potential health research impact questions (e.g., “Does our health services research program on decreasing wait times have an important impact on policy makers and health care?”). Currently, 2 indicators from the decision-making category relate to policy-maker impacts and a selection from the determinants of health subcategory on accessibility could be used to track these impacts.

Given the scarcity of funding for health research in Canada, strategic selection of indicators that are sensitive and specific enough to address evaluation questions, while not being too expensive or too time-consuming collect, will be a major challenge. Only the most important return-on-investment questions can realistically be considered.

Conclusions and the way forward

Both econometric and payback methods have demonstrated that economic returns from health research have been sizeable and are ongoing. Although these assessments support the notion that some health research investment is worthwhile, the latest comprehensive assessment of this topic in the United Kingdom has just confirmed that a standardized and mapped way of classifying research funding with a deeper understanding of international flows of knowledge and influence is still required.

The Canadian Academy of Health Sciences has developed a new method to meet these needs. It proposes a new framework and a library of preferred indicators (the individual strength and specificity of which are critical) that are to be selected in customized sets to address specific impact evaluation questions from a variety of health research funders and stakeholders.

This approach will not be fast, easy or inexpensive to apply, and substantial collaboration is recommended in implementing it. However, the approach should help optimize limited resources for the evaluation of impact, eliminate duplication and double-counting of impacts, and help offset the costs of evaluating impacts this well. The use of this new model and its approach to understand and progressively enhance the impact of health research investments will indirectly contribute to improved health, economic and social outcomes for all Canadians.

Competing interests: None declared.

Contributors: Both of the authors contributed to the conception and writing of the manuscript. Both revised it for intellectual content and approved the version submitted for publication.

Acknowledgements: The authors thank the sponsors of the major assessment for their financial support of the review that allowed this analysis, the individual strength and specificity of which are critical. Presented at the Canadian Academy of Health Sciences Annual meeting, 18 Sept. 2007. 2009. p. A42-74. Available: www.cahs-acss.ca/e/assessments/completedprojects.jsp (accessed 2009 Jan. 21).

Access Economics. Exceptional returns: the value of investing in health R&D in Australia. Canberra (Australia): Australian Society for Medical Research; 2003. Available: www.accessconomics.com.au/publications/reports/getreport/report=33&id=40 (accessed 2009 Jan. 21).

Access Economics. Exceptional returns: the value of investing in health R&D in Australasia II. Canberra (Australia): Australian Society for Medical Research; 2008. Available: www.amer.org.au/EExceptionalReturns2008.pdf (accessed 2009 Jan. 21).

Buxton MJ, Hanney S, Hyman S, Jones T. Estimating the economic value to societies of the impact of health research: a critical review. Bull World Health Organ. 2004;82(7):733-9. Available: www.bruel.ac.uk/3269/Herg/17373.pdf (accessed 2009 Jan. 21).

Buxton M, Hanney S. Development of a multidimensional ‘payback’ framework for assessing investments in health research. Ital J Public Health. 2005;2:60.

Wooding S, Hanney S, Buxton M, et al. Payback arising from research funding: evaluation of the Arthritis Research Campaign. Rheumatol (Oxford). 2004;14:1145-56.

Developing a CIHR framework to measure the impact of health research. Ottawa (ON): Canadian Institutes of Health Research; 2005. Available: www.cihr-irsc.gc.ca/e/documents/meeting_synthesis_e.pdf (accessed 2009 Jan. 7).

Buxton M, Hanney S, Morris S, et al. Medical research – what’s it worth? Estimating the economic benefits from medical research in the UK. London (UK): Health Economics Research Group, Office of Health Economics, and RAND Europe; 2008. Available: www.acmedsci.ac.uk/rpp99uid137.html (accessed 2009 Jan. 8).

Adson JM, Pardey PG. Attribution and other problems in assessing the returns to agricultural R&D Agriculture Economics. 2001;25:141-52.

Canadian Academy of Health Sciences. Making an impact: A preferred framework and indicators to measure returns on investment in health research. Ottawa (ON): The Academy; 2009. Available: www.cahs-acss.ca/e/assessments/completedprojects.jsp (accessed 2009 Jan. 21).

Schechter MT, Armstrong PW. The Canadian Academy of Health Sciences: coming of age at the right time. CMAJ 2008;178:1029-31. Available: www.cahs-acss.ca/e/assessments/completedprojects.jsp (accessed 2009 Jan. 8).

Prospectus for a major assessment: the return on investments in health research: defining the best metrics. Ottawa (ON): Canadian Academy of Health Sciences; 2007. Available: www.cahs-acss.ca/e/docs/ROI_Prospectus.pdf (accessed 2009 Jan. 8).

Brutscher P-B, Wooding S, Grant J. RAND Europe. 2008. Health research evaluation frameworks — an international comparison. Cambridge (UK): RAND Corporation; 2008. Available: www.rand.org/pubs/technical_reports/2008/RAND_TR629.pdf (accessed 2009 Jan. 8.)

Correspondence to: Dr. Cy Frank, Department of Surgery, University of Calgary, Rm. 451B-3330, Hospital Dr. NW, Calgary AB T2N 4N1; fax 403 283-7742; cfrank@ucalgary.ca