

This is an electronic version of an article published in *Research Evaluation*, 19(4), October 2010, pages 251-261.

<http://www.ingentaconnect.com/content/beechn/rev/2010/00000019/00000004/art00003?crawler=true>

**Evaluation of  
Research, Technology and Development:  
From “Prescriptions for Justifying”  
To “User-Oriented Guidance for Learning”**

**Steve Montague  
Performance Management Network Inc.  
Ottawa, Ontario, Canada  
[steve.montague@pmn.net](mailto:steve.montague@pmn.net)**

**Rodolfo Valentim  
Canadian Cancer Society Research Institute  
Toronto, Ontario, Canada  
[rvalentim@cancer.ca](mailto:rvalentim@cancer.ca)**

## **Abstract**

The measurement and evaluation of research, technology and development (RT&D) has gone through phases over the past 50 years. Over time, high-level measures such as total expenditures on R&D, overall citations and patent production have given way to more contextualized metrics recognizing the inherent differences in innovation subject areas and the need to show mission achievement. This article shows how recently proposed Canadian Academy of Health Sciences (CAHS) metrics were adapted to help frame a case study conducted by the Canadian Cancer Society Research Institute (CCSRI). Early results suggest that the framework provides a useful structure to display both a hierarchy of results focused on mission goals, and to build an attributable RT&D and innovation story over time. With this work and other recent developments, evaluation appears poised to go beyond retrospective justification and to become a fully legitimate part of strategic learning for RT&D initiatives.

## The Situation

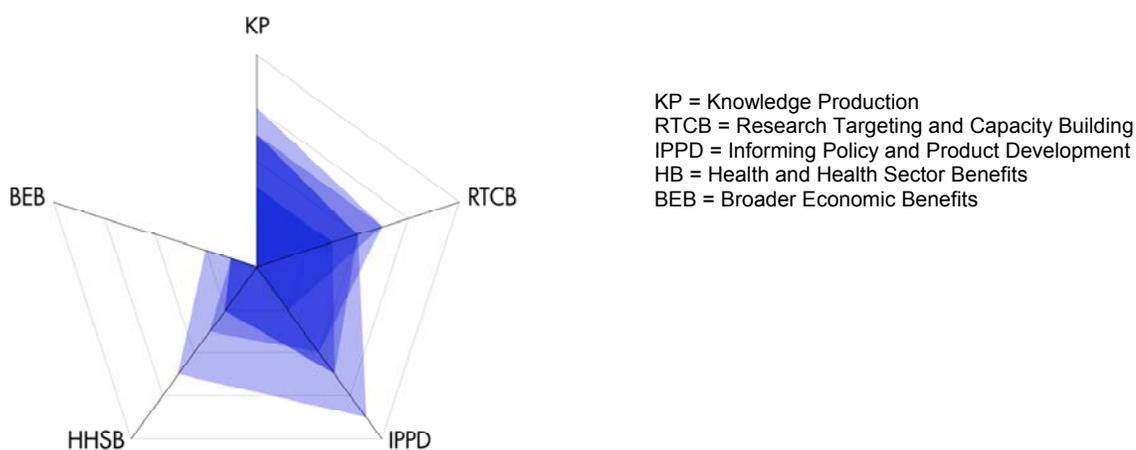
The evaluation community has struggled to address the performance of research, technology and development (RT&D) (also known as research and development or the broader R&D and innovation) for decades. Traditionally the 'prescription' of measures for RT&D has involved measurement of the following:

- ▶ inputs (money spent, occasionally monetizing the value of all human resource inputs)
- ▶ outputs (the tangible 'products' of research - typically papers, presentations, etc.)
- ▶ outcomes (defined as publications<sup>1</sup> and publication 'productivity', citations)
- ▶ impacts (generally viewed as the use of research for the benefit of society)

Gault (2006) reviews the history of R&D and innovation statistics over past decades. Initially, input indicators such as Gross Expenditures on R&D (GERD), investments in R&D personnel and the like were measured and "assumed" to result in improved innovation. Such efforts were assisted by the Frascati manual, which was first drafted in 1962. The Frascati manual "deals exclusively with the measurement of human and financial resources devoted to Research and Experimental Development" (OECD 1992). Policy discussions of the early 1980s certainly featured inputs such as these as key markers of progress. These indicators have persisted through the decades. As Shodjai (1996) explains, "in the most prevalent structures of S&T activities, the units of measurement are money and people." In the same era, system 'output' indicators related to intellectual property production (publications and patents, etc.) and highly qualified personnel were also counted. There was often an assumed link to innovation and economic growth.

Economic outcomes in the pre-1990s era were often somewhat theoretical using broad assumptions related to factors such as the generation of consumer or supplier surplus (Mansfield *et al.*, 1977). In this approach, the social benefits from an innovation are measured by the profits of the innovator from the innovation plus the benefits to consumers following a reduction in the price of the goods or services due to the innovation. A main limitation of his model was that it captured only (immediately) monetized market effects for single products or processes. Knowledge or network effects were generally not covered.

More recently, starting in the 1990s and moving into the 2000s we have seen scorecard and 'payback' approaches emerge (Wooding *et al.*, 2004). These models show categories of impact and either precisely (i.e. they use actual metrics) or imprecisely (i.e. they use expert rating) provide scores which can be depicted in tabular form or in the classic spider diagram as in Figure 1.



**Figure 1. Visual representation of 'payback profiles'**

Median payback scores are plotted on one axis of a spider diagram whose size and shape correspond to the payback from a particular case study. Case studies can be compared by overlaying the profiles of a group of case studies (Wooding *et al.*, 2004).

Metrics have tended to run across a spectrum. The use of aggregate expenditures and economic statistics like Government expenditure on R&D and the overall market share for certain innovative products has been favoured by macro policy makers (e.g. the Frascati conventions in 1962 and several National Government summaries since then). Micro economic cost benefit analyses have been favoured by some S&T economists dealing with micro economic policies (e.g. see Mansfield 1977 and beyond). Traditional measures of scientific merit such as publications and citations have been favoured by academia. All of these historical evaluation approaches have their relative merits and weaknesses. A summary assessment is contained in Table 1.

**Table 1. Critique of traditional research, technology and development evaluation approaches**

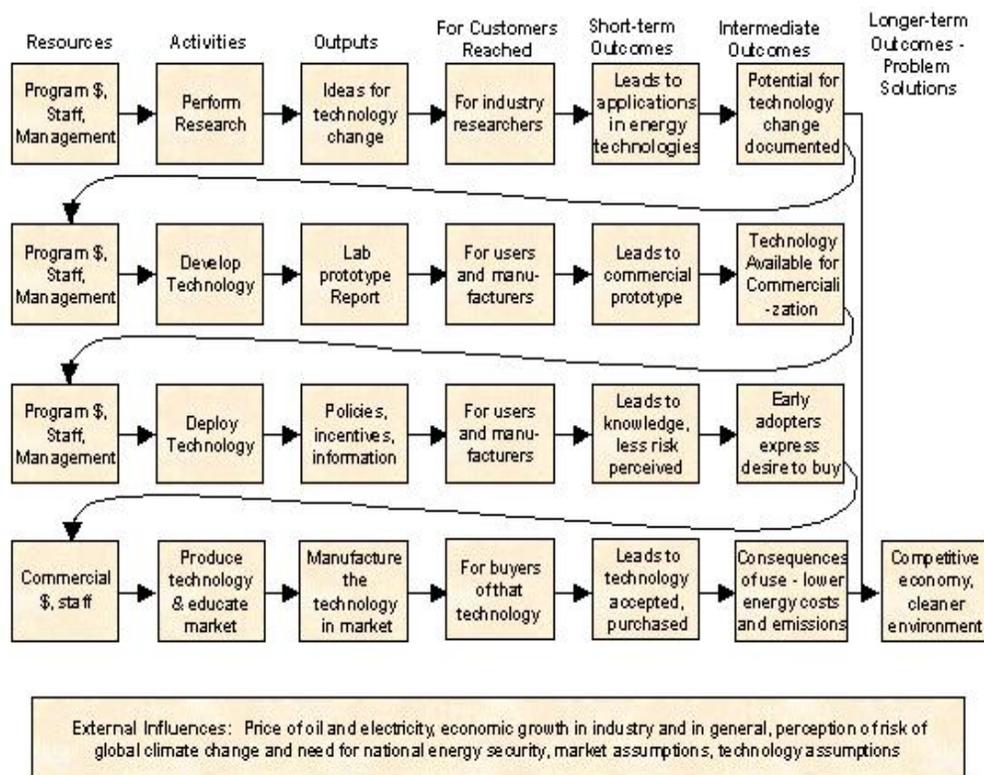
Examples	Assessment
Aggregate Economic Statistics - Expenditures on R&D - GERD - Market share - Trade deficit in innovative goods	<ul style="list-style-type: none"> <li>▪ The measures are not 'outcomes' <i>per se</i></li> <li>▪ Attribution problems – too high level</li> <li>▪ Understanding?</li> </ul>
Micro-economic Benefit-Cost - Benefit-Cost Analysis - Cost-Effectiveness Analysis	<ul style="list-style-type: none"> <li>▪ Favours easily quantifiable 'outcomes'</li> <li>▪ Monetization problems</li> <li>▪ High sensitivity to assumptions</li> </ul>
Scientific Merit Valuation - Peer review - Publications - Citations	<ul style="list-style-type: none"> <li>▪ Partial 'outcomes'</li> <li>▪ Bias problems – science viewpoint (self citations, citation 'circles' and citation context are all threats to validity)</li> <li>▪ Misses the bigger system</li> </ul>

In addition to the weaknesses present in each of these approaches, there is a fundamental issue regarding the interpretation of performance. The problem with metrics being used as 'prescriptions' or standardised 'benchmarks' is that they:

1. Impose metrics which may or may not fit the nature of the innovation area. For example, some areas of R&D and innovation emphasize publishing more than others. This means that publication productivity and citation metrics will favour certain areas of science over others. The same can be said for the filing of intellectual property (IP) protection through patents, where some areas favour formalized IP over others. Ultimately, metrics need to be developed in context.
2. Fail to tell a performance story relating R&D and innovation to impacts and benefit. In addition to context, the understanding of the merit of a given investment in science or innovation depends on understanding its place in a sequence or more likely a cluster (or clusters) of activity focussed in given areas. See Jordan "A Theory-Based Logic Model for Innovation Policy and Evaluation" in this same issue. One must understand the causal connections (potential or actual) between areas. Scorecards or dashboards do not typically make those connections.
3. Imply a supply-side science, linear model. Related to the points above, analysts like Shodjai (1996), Rosenberg (1991) and Kuhlmann (2001) have noted the persistence of a linear model of innovation in which investments lead to research, which lead to discovery and invention, which move on to innovation and diffusion. This has been widely agreed as being too simple minded, yet many current depictions of innovation programs keep showing it. As Rosenberg states, the linear model "is dead but won't lie down" (Shodjai, 1996).

## Recent Developments

Recent developments from the mid 1990s onward have focussed on the use of logic models to tell a RT&D performance story (McLaughlin & Jordan, 1999). These were initially used for micro-level projects, but could also be applied at higher levels. Figure 2 depicts an example of how a logic model can be used to illustrate the logical flow and linkages that are part of a performance story. Note that while the model follows a linear pattern 'reach' is now shown, which recognizes different players in the innovation system.



**Figure 2. Logic Chart for a Research and Technology Development and Deployment Program**

The logic model is intended to show the logical flow of resources and activities leading to outputs to intermediate outcomes and long-term outcomes or impacts. Inputs are placed in the first column on the left and longer-term outcomes on the far right (McLaughlin and Jordan, 1999).

Kuhlmann (2001) notes the further development of RT&D and S&T innovation thinking in terms of trends such as the 'soft side' of innovation (i.e. the learning ability of all actors in the innovation process) and what Gibbons et al (1994) call Mode-2 science. Mode-1 refers to traditional science-driven modes of knowledge production. Mode-2 refers to knowledge production processes stimulated and influenced far more by demand. In this mode many actors other than scientists have important and recognized roles to play.

This kind of cause and effect thinking was also applied in a slightly more 'open' format in Canadian work during the 1990s (Teather and Montague, 1997). In one approach, the sequence of effects can be shown more as a checklist, or set of touchstones, as represented in Table 2. This template attempts to integrate a wide range of impacts into a structured quantitative and qualitative assessment (Teather and Montague, 1997).

**Table 2. A Tabular Example of Cause and Effect Thinking**

<b>Case Impacts – A Structural Assessment</b>			
<b>Project Incrementality (Direct Influence)</b>	<b>Direct Collaborator / Performer / User Impacts</b>	<b>Industry Sector / Supply Community Impacts</b>	<b>Economy / Societal Impacts</b>
Involvement <input type="checkbox"/> project would not have been done without program assistance <input type="checkbox"/> contributed to completing the project more quickly <input type="checkbox"/> contributed to completing the project more thoroughly	Technical results <input type="checkbox"/> new or improved product <input type="checkbox"/> new or improved process <input type="checkbox"/> advancement of knowledge <input type="checkbox"/> increased technical capabilities <input type="checkbox"/> improved quality control <input type="checkbox"/> new skills internally <input type="checkbox"/> increased efficiency / improved productivity <input type="checkbox"/> technology transfer Policy / legislative results <input type="checkbox"/> policy behavioural changes <input type="checkbox"/> agreement / accord <input type="checkbox"/> legislation / regulation Infratechnology results <input type="checkbox"/> Codes, standards, databases, protocols <input type="checkbox"/> acceptance of standards Commercial results <input type="checkbox"/> increased sales <input type="checkbox"/> increased market share <input type="checkbox"/> increased profitability <input type="checkbox"/> cost savings Organizational effects <input type="checkbox"/> increase in jobs <input type="checkbox"/> diversification <input type="checkbox"/> expansions <input type="checkbox"/> strategic alliances / partnerships <input type="checkbox"/> achievement awards / recognition	<input type="checkbox"/> production process efficiencies <input type="checkbox"/> increased science and technology information <input type="checkbox"/> increased sales <input type="checkbox"/> cost savings <input type="checkbox"/> changes to industry structure (e.g., concentration, competitiveness internationally) <input type="checkbox"/> spin-off companies <input type="checkbox"/> technology infrastructure (e.g., standard scientific and engineering data, industry standards, test protocols, and instrumentation) <input type="checkbox"/> training of technological problem-solvers whose talents can be applied in many areas <input type="checkbox"/> establishment of quality, performance standards	Economic Benefits <input type="checkbox"/> reduced consumer costs <input type="checkbox"/> increased employment <input type="checkbox"/> improved competitiveness <input type="checkbox"/> reduction in subsidies Societal Benefits <input type="checkbox"/> improved quality of life <input type="checkbox"/> protection of environment <input type="checkbox"/> improved energy efficiency <input type="checkbox"/> improved public health and safety <input type="checkbox"/> education / awareness <input type="checkbox"/> public service efficiency gains (i.e. lowered taxpayer burden)

Finally, innovation diffusion theories (Rogers, 1995 and Reed and Jordan, 2007) and more recently knowledge translation perspectives (Davison, 2009) have led to the establishment of generalized results logic models such as the Canadian Academy of Health Sciences (CAHS) model depicted in Figure 3. Note that the chart moves from left to right in an apparently 'linear' fashion, although it distinctively shows the "back-and-forth" interaction among key stakeholder groups and the building of a pool of knowledge.

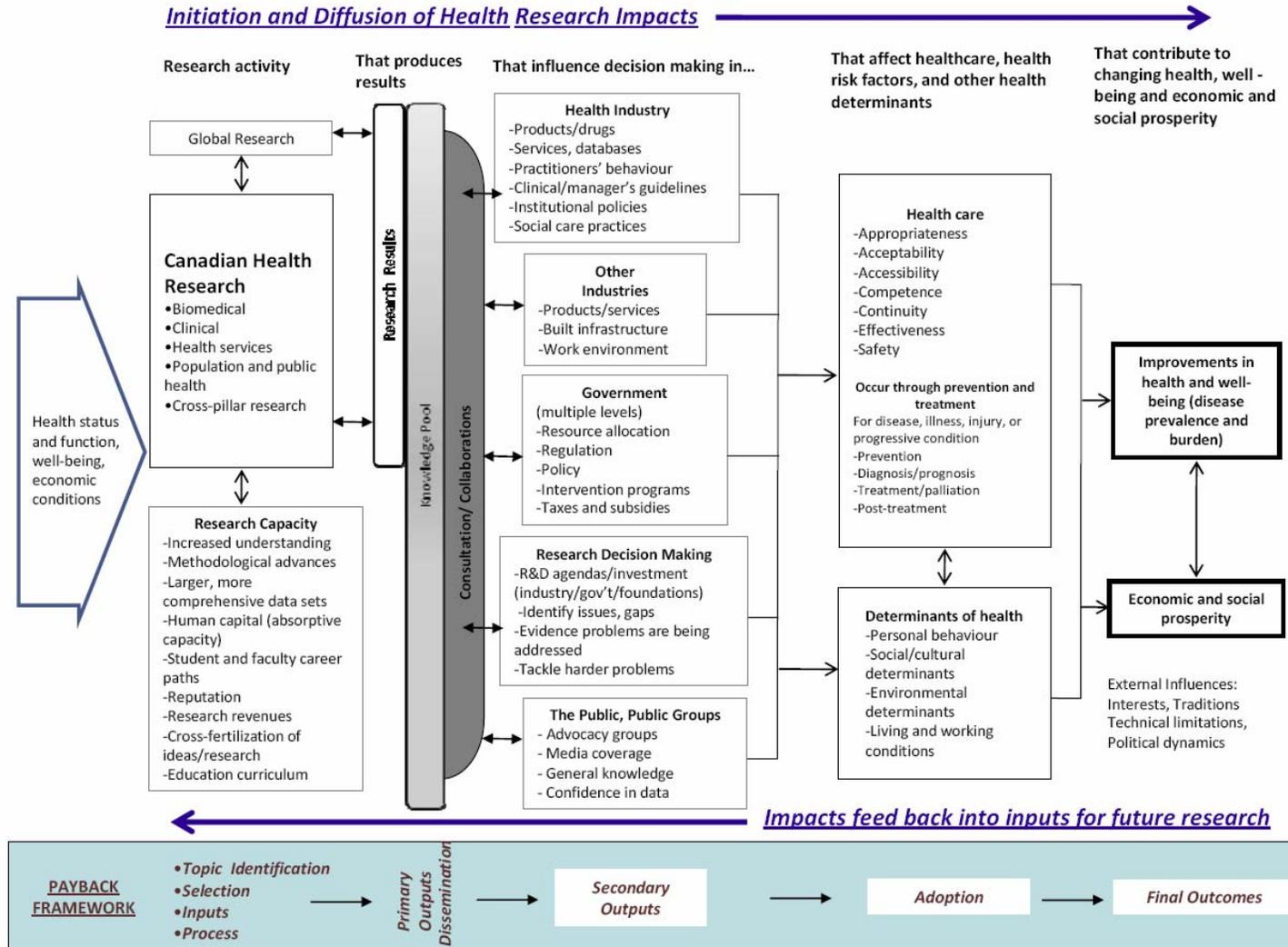


Figure 3. CAHS Framework Logic Model of Health Research Progression to Impacts

The framework is read from left to right and attempts to illustrate how research activity influences decision making, which results in health improvements, as well as economic and social prosperity. The framework is designed to track health research impacts in five categories: 1. advancing knowledge, 2. building capacity, 3. informing decision-making, 4. health impacts, 5. broad socio-economic impacts (Panel on Return on Investment in Health Research, 2009).

CAHS suggests 60 indicators within 5 categories of impact. These are broadly described in Table 3. Note that while they are similar to the categories displayed in Figure 1, when related to a results logic these may show a kind of innovation progression.

**Table 3. Describing Canadian Academy of Health Sciences Impact Categories**

<ul style="list-style-type: none"> <li>▶ Advancing knowledge indicators and metrics include measures of research quality, activity, outreach and structure. Identified are 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).</li> <li>▶ 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.</li> <li>▶ 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. Provided are two aspirational indicators for this category (media citation analysis and citation in public policy documents).</li> <li>▶ 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.</li> <li>▶ Broad economic and social impacts are classified into activity, commercialization, health benefit (specific costs of implementing research findings in the broad health system), wellbeing, and social-benefit indicators (socio-economic benefits).</li> </ul>
---

Each impact category consists of subcategories, which identify evaluation methods and data that facilitate impact evaluation and identification of contributing factors (Panel on Return on Investment in Health Research, 2009).

### **Going Further**

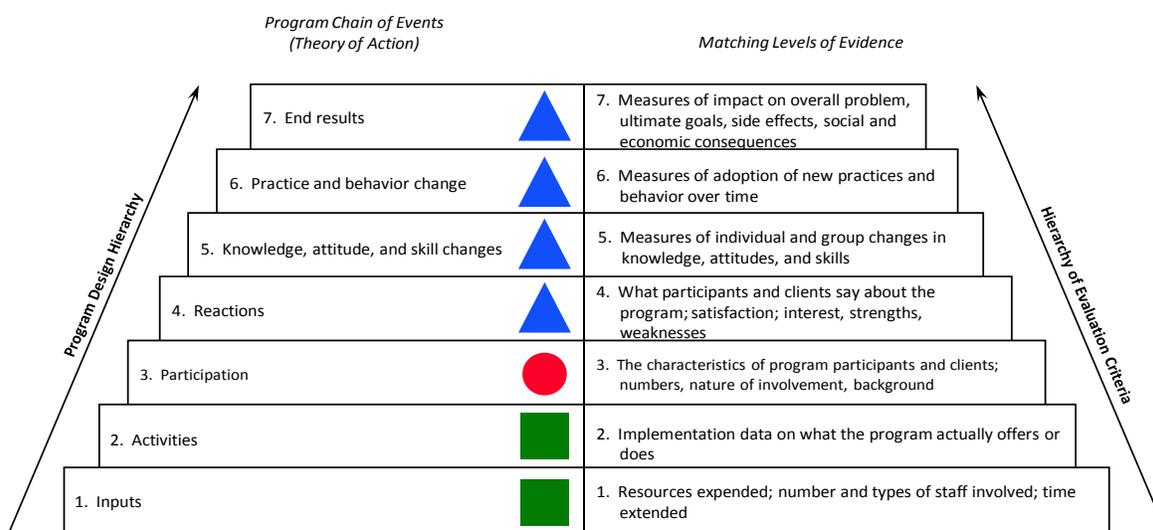
The question regarding this framework is how can it be put into practical use? Past decades have shown that, despite its obvious short-comings, linear 'supply side' orientations to measuring RT&D and innovation progress persist. It would appear that a simple linear looking model, while flawed, has tended to be favoured over less 'directive' models. How can we shift the focus to reach, usage and interactive systems changes without over-complicating the approach?

Recent developments suggest that there is a trend away from simply 'justifying' R&D investments, using simplified metrics and using analytical methods that provide little to no insight on how innovation works. Emerging is an era in which R&D and innovation can be analysed using theories of change, innovation diffusion, knowledge transfer and translation. The emphasis is moving from linear input-output-outcome thinking to the inclusion of reach and the behavioural influence of innovations on key groups and systems. The question now is can we establish a straight-forward approach to help us in this effort?

We think we have done this in recent work with the Canadian Cancer Society. The approach is presented in the following case example.

### The Canadian Cancer Society: A Results Chain Hierarchy and the MA 17 Clinical Trial Case Study

Over the past three years, the Canadian Cancer Society (the Society), Canada's largest health charity, has been implementing an approach to results planning, measurement, evaluation and reporting which has been based on what has become known as a 'Bennett Hierarchy' (Bennett and Rockwell, 1995). Figure 4 shows this diagram. Note that the notion of "reach" is present in Step 3 "Participation".



Source: Adapted from Claude Bennett 1979. Taken from Michael Quinn Patton, *Utilization-Focused Evaluation: The New Century Text*, Thousand Oaks, California, 1997, p 235.

### Figure 4. Theory of Action: Levels of Evidence and the Logic Model

Source: Adapted from Bennett (1979). Taken from Patton (1997: 235)

The hierarchical model shows the seven levels of events / results on the left hand side and the matching level of evidence on the right hand side.

The results chain hierarchy has proven useful for Society planning and reporting in areas such as direct service delivery and advocacy. The approach has been valuable both in terms of planning and reporting progress for accountability, and as a learning tool. In the Spring of 2009 the Canadian Cancer Society Research Institute decided to adapt and apply the hierarchy in the assessment of the impact of funded research. In doing so, the indicators suggested by CAHS (Panel on Return on Investment in Health

Research, 2009) were sorted according to the results chain hierarchy. Figure 5 illustrates the summary reconciliation of this work.

Initiative Chain of Results	Hierarchy of Evaluation Criteria / Evidence	Typical Indicators	Typical Sources / Methods
End outcomes 	Measures of impact on overall problem, ultimate goals, side effects, social and economic consequences	<ul style="list-style-type: none"> <li>• Rate or incidence of cancer (incidence, mortality, morbidity)</li> <li>• Level of quality of life (Index TBD)</li> <li>• Level of advances in cancer science / research</li> </ul>	<ul style="list-style-type: none"> <li>• Specialized analyses / evaluations</li> <li>• Statistical agency data</li> <li>• Canadian cancer statistics</li> <li>• Analytical and policy groups</li> </ul>
Practice and behaviour change 	Measures of adoption of new practices and behaviour over time	<ul style="list-style-type: none"> <li>• Level of research used (knowledge transfer, practice adoption) by scientists / policy makers / institutions / health care practitioners / consumers</li> <li>• Level of research used in curricula for new researchers (citation in text books and reading lists)</li> <li>• Level of research cited in ongoing health professional education material</li> <li>• Level of research cited in public policy documents</li> <li>• Level of research cited in advocacy publications</li> </ul>	<ul style="list-style-type: none"> <li>• Physical observation</li> <li>• Inspections, reviews</li> <li>• Surveys</li> <li>• Evaluation studies</li> <li>• Content analysis</li> <li>• Bibliometrics (citation analysis)</li> </ul>
Knowledge, attitude, skill and aspiration change 	Measures of individual and group changes in knowledge, abilities, skills and aspirations	<ul style="list-style-type: none"> <li>• Level of understanding of key related science information generated through research by scientists / policy makers / institutions / health care practitioners / consumers</li> <li>• Level of self-expressed commitment to specific areas of science / research or practice / protocol / policy change by scientists / policy makers / institutions / health care practitioners / consumers</li> <li>• Level of development of new knowledge in cancer research</li> <li>• Level of development of new methods in cancer research</li> <li>• Level of published research findings in a timely manner and in peer-reviewed journals with high "impact factors"</li> </ul>	<ul style="list-style-type: none"> <li>• Independent review of target groups</li> <li>• Content analysis</li> <li>• Survey, group self-assessment</li> <li>• Testing / certification</li> <li>• Bibliometrics (publications)</li> </ul>
Reactions 	What participants and clients say about the program; satisfaction; interest, strengths, and weaknesses	<ul style="list-style-type: none"> <li>• Level of program recognition and support from key stakeholders / target groups / participants</li> <li>• Level (volume, accuracy and 'tone') of media coverage of research and program activities</li> </ul>	<ul style="list-style-type: none"> <li>• Usage / participation tracking</li> <li>• Correspondence content analysis</li> <li>• Surveys</li> <li>• Media content analysis</li> </ul>
Engagement / participation 	The characteristics of program participants and clients; number, nature of involvement, and background	<ul style="list-style-type: none"> <li>• Level of engagement with other centres, networks, academic institutions, government agencies, etc.</li> <li>• Level of engagement by stakeholders / target groups / participants</li> <li>• Level of multidisciplinary and / or multisectorial research activities</li> <li>• Level of recruitment and retention of stakeholders / target groups / participants (e.g. junior investigators, researchers, review panellists, etc.)</li> <li>• Level of established external scientific advisory board(s)</li> </ul>	<ul style="list-style-type: none"> <li>• Web use tracking</li> <li>• Correspondence content analysis</li> <li>• Observation of meetings / events</li> <li>• Meeting attendance records</li> <li>• Stakeholder relationship management / tracking (e.g. contracts and agreements)</li> <li>• Surveys</li> </ul>
Activities & outputs 	Implementation data on what the program actually offers	<ul style="list-style-type: none"> <li>• Level of research as per internal review guidelines</li> <li>• Extent to which plans, strategies, frameworks, etc. are delivered as per expectations (expected timelines, resource usage and quality levels)</li> <li>• Extent to which governance structure adheres to internal guidelines</li> <li>• Extent to which policy and financial decisions are made according to Board / Senior Management / Expert Advisory Committee(s) accepted guidelines and standards</li> <li>• Extent to which internal and external communication strategies adhere to internal standards and protocols / policies</li> </ul>	<ul style="list-style-type: none"> <li>• Project / initiative tracking</li> <li>• Project reports</li> <li>• Content analysis or records</li> <li>• Peer-review</li> <li>• Operating reviews</li> </ul>
Inputs 	Resources expended; number and types of staff involved; time expended	<ul style="list-style-type: none"> <li>• Level of human resources (staffing) at all levels (according to norms, vacancies, expectations, benchmarks)</li> <li>• Level of financial resources (budgets vs. actuals) at all levels</li> </ul>	<ul style="list-style-type: none"> <li>• Budget analysis</li> <li>• Time, reporting and budget / plan review</li> <li>• Activity-based costing</li> </ul>

Figure 5. A Results Chain Typical Indicator and Measurement “Menu” for CCSRI National Programs and Networks

The results chain hierarchy shows the relationship among the resources that are invested, the activities that take place, and the sequence of changes that result. Figure 5 should be read from the bottom up and from left to right. The first column identifies the levels of results (1 – 7); the second provides a description of the levels of results; the third provides a set of generic indicators for each of the levels of results; and the fourth lists typical sources of data and methods of evaluation. This chart was developed as a starting 'menu' for the description of Society RT&D cases.

The use of the results chain hierarchy to evaluate research impacts is still in its early days of implementation, however the framework was recently applied to develop an evaluative performance story for a clinical trial for letrozole, an aromatase inhibitor used as a late adjuvant therapy for hormone receptor positive (HR+) breast cancer. A summary of the case study, key study findings and the performance story is described below.

Up until the early 2000s, the standard for adjuvant treatment of women with early stage, HR+ breast cancer was the prescription of a drug named tamoxifen. A selective estrogen receptor modulator, tamoxifen works by binding to the estrogen receptor thus keeping cells from making or using estrogen. While tamoxifen was shown to be effective at reducing cancer recurrence risk for the first few years after surgery, research demonstrated that continuing tamoxifen therapy for longer than five years did not improve recurrence-free or overall survival and could actually be harmful (Burstein, 2003). There were no widely accepted treatments available beyond five years, yet research showed that a significant risk of recurrence for breast cancer still remained. Toronto oncologist and cancer researcher Dr. Paul Goss, currently of the Massachusetts General Hospital Cancer Center, had preliminary research findings which suggested that the administration of an aromatase inhibitor, a class of drugs that, unlike tamoxifen, actually suppress the synthesis of estrogen, could be an effective adjuvant therapy for postmenopausal women with HR+ breast cancer. He approached the Canadian Cancer Society funded NCIC Clinical Trials Group (CTG), a cooperative oncology network based at Queen's University, with the idea for a clinical trial using an aromatase inhibitor sequentially with tamoxifen to reduce breast cancer recurrence in post menopausal women with HR+ breast cancer.

Dr. Goss eventually led a group that would test this hypothesis by conducting the MA 17 clinical trial. The MA 17 clinical trial was activated in August 24, 1998 and met patient accrual by mid-2002. The trial was organized and coordinated across two continents, six distinct institutional groups, nine countries and 416 sites by the CTG. An interim analysis of the planned five year study demonstrated results that were so convincing that CTG leadership, in consultation with the data and safety monitoring committee (DSMC), decided to terminate the trial early and unblind the study in October, 2003 in order to disseminate the remarkable preliminary results. In November 2003, CTG published the findings in the *New England Journal of Medicine*, showing that the use of the aromatase inhibitor letrozole significantly reduced the risk of breast cancer recurrence in the late adjuvant setting. This led to a virtually immediate and traceable change in treatment practices.

The MA 17 trial advanced the availability of an effective therapy in a clinical setting where there had not been one before. Under the strong leadership of principal investigator Dr. Goss, and with the engaged support of a diverse set of researchers, the CTG delivered persuasive evidence that has had a global impact on preventing disease recurrence in women surviving breast cancer. This reach was facilitated by the ability of the CTG to attract a strong group of prestigious international researchers and institutions to the trial. Upon study unblinding, there was a widely distributed press release, followed by a number of highly-cited publications. Post trial market research conducted by the pharmaceutical company that participated in the trial suggested that prescribing physicians recalled the name of the trial (MA 17), the name of the principal investigator (Dr. Goss) and the coordinating centre of the trial (CTG) in an essentially unprecedented fashion, suggesting that the trial directly influenced their prescribing patterns.

Critical factors in advancing this very successful clinical trial were the broad network structure of the CTG, including loose ties to several key groups and a potentially low perceived bias towards any given group or approach, the longstanding quality of work and reputation of the CTG Breast Disease Site Committee and the drive of the principal investigator Dr. Goss. Also worth noting is the Society's long history of funding hormone base breast cancer research, which, although not established as fundamental to the trial itself,

was part of the foundational endocrine-related breast cancer research that preceded and undoubtedly informed the trial. This point highlights the value of funding discovery-based cancer research as it contributes to the knowledge base and research capacity that may ultimately inform and lead to important breakthroughs further down the research and development (R&D) and innovation spectrum.

### **Outcomes and Impacts of the MA 17 Clinical Trial**

The outcomes and impacts of the MA 17 trial are summarized in the results hierarchy shown in Figure 6 and include the following observations:

- ▶ Level 3 (engagement) – The MA 17 trial engaged a large and prestigious international group of researchers. This international engagement has fostered improved communication across groups and has heightened the CTG's status in the international clinical trials research community. It has also increased the CTG's ability to develop, mobilize and coordinate large international trials.
- ▶ Level 4 (reactions) – As a result of a high level of engagement, the CTG was invited to become a member of the US NCI clinical trials cooperative group program. There was also unprecedented media coverage of trial results.
- ▶ Level 5 (knowledge, attitude, skill and aspiration change) – New knowledge was generated about the role of aromatase inhibitors in adjuvant therapy, as evidenced through the publication list of the MA 17 trial. Nineteen publications in high impact journals and 28 abstracts were presented or published. Additionally, there was unprecedented recognition of trial performers (MA 17, Dr. Goss, CTG) by prescribing physicians.
- ▶ Level 5 (knowledge, attitude, skill and aspiration change) – The MA 17 trial showed the value of pursuing adjuvant breast cancer therapy beyond five years. This indicates a paradigm shift for research and clinical treatment.
- ▶ Level 6 (practice and behaviour change) – The primary trial results publication (Goss et al., 2003) is cited 721 times according to the Web of Science database. This high citation count indicates that trial results have informed a new research direction in late adjuvant therapy.
- ▶ Level 6 (practice and behaviour change) – The MA 17 trial was a success, demonstrating the value of aromatase inhibitors in treating postmenopausal women with HR+ early-stage breast cancer in the adjuvant setting. It established an aromatase inhibitor, letrozole, as a standard adjuvant therapy for breast cancer, following tamoxifen treatment. The trial of letrozole as an extended adjuvant therapy had essentially no 'competitor' technologies. (While there were certainly other aromatase inhibitors in the marketplace – there were none being used for late adjuvant therapy.) A similar trial was started three years after MA 17, suggesting at least a three year lead over any similar work.
- ▶ Level 6 (practice and behaviour change) – The results of the MA 17 trial contributed to changing world-wide standard of care practices and are cited prominently in provincial, national and international guidelines for the adjuvant treatment of breast cancer. Based on trial results, adjuvant therapy for postmenopausal women with HR+ breast cancer is suggested to include an aromatase inhibitor in order to lower the risk of tumour recurrence. More specifically, Cancer Care Ontario, the British Columbia Cancer Agency, the Alberta Cancer Board, the National Comprehensive Cancer Network, and the American Society of Clinical Oncology guidelines all recommend aromatase inhibitors for women with HR+ breast cancer. The MA 17 trial is specifically cited in the recommended clinical practice guidelines of all these groups.
- ▶ Level 7 (end outcomes) – Evidence suggests that the MA 17 clinical trial was responsible for hundreds and possibly thousands of disease free years for Canadian women. It has almost certainly created thousands of disease free years worldwide.

Chain of Results	Hierarchy of Evaluation Criteria / Evidence	Typical Indicators	Outcomes and Impacts
 7. End Outcomes	Measures of impact on overall problem, ultimate goals, side effects, social and economic consequences	<ul style="list-style-type: none"> <li>✓ Rate or incidence of cancer (incidence, mortality, morbidity)</li> <li>– Level of quality of life (Index TBD)</li> <li>✓ Level of advances in cancer science / research</li> </ul>	<ul style="list-style-type: none"> <li>▪ Reduced recurrence of breast cancer, which resulted in hundreds of disease free years for Canadian women</li> <li>▪ Increased use of aromatase inhibitors in adjuvant therapy</li> <li>▪ Increased international collaboration on hormone based breast cancer research</li> </ul>
 6. Practice & Behaviour Change	Measures of adoption of new practices and behaviour over time	<ul style="list-style-type: none"> <li>✓ Level of research used (knowledge transfer, practice adoption) by scientists / policy makers / institutions / health care practitioners / consumers</li> <li>– Level of research used in curricula for new researchers (citation in text books and reading lists)</li> <li>✓ Level of research cited in ongoing health professional education material</li> <li>✓ Level of research cited in public policy documents</li> <li>✓ Level of research cited in advocacy publications</li> </ul>	<ul style="list-style-type: none"> <li>▪ Adherence to new protocols by practitioners regarding extended adjuvant therapy options</li> <li>▪ Changed timeframe for adjuvant therapy research</li> <li>▪ Increased number of clinical trial using aromatase inhibitors in the adjuvant setting</li> <li>▪ New standard for adjuvant therapy (use of aromatase inhibitors and longer term adjuvant treatment)</li> <li>▪ Trial results cited in provincial, national and international clinical practice guidelines</li> <li>▪ Primary trial publication (Goss, 2009) cited 721 times (WOS, 2010)</li> </ul>
 5. Knowledge, Attitude, Skill and Aspiration Change	Measures of individual and group changes in knowledge, abilities, skills and aspirations	<ul style="list-style-type: none"> <li>✓ Level of understanding of key related science information generated through research by scientists / policy makers / institutions / health care practitioners / consumers</li> <li>✓ Level of self-expressed commitment to specific areas of science / research or practice / protocol / policy change by scientists / policy makers / institutions / health care practitioners / consumers</li> <li>✓ Level of development of new knowledge in cancer research</li> <li>✓ Level of development of new methods in cancer research</li> <li>✓ Level of published research findings in a timely manner and in peer-reviewed journals with high “impact factors”</li> </ul>	<ul style="list-style-type: none"> <li>▪ Increased interest in late adjuvant therapy research</li> <li>▪ Demonstrated value of pursuing adjuvant breast cancer therapy beyond five years</li> <li>▪ Unprecedented recognition of trial performers (MA 17, Dr. Goss, CTG) by prescribing physicians</li> <li>▪ 19 publications in high impact journals including the New Journal of Medicine and Journal of Clinical Oncology; 28 abstracts presented or published</li> </ul>
 4. Reactions	What participants and clients say about the program; satisfaction; interest, strengths, and weaknesses	<ul style="list-style-type: none"> <li>✓ Level of program recognition and support from key stakeholders / target groups / participants</li> <li>✓ Level (volume, accuracy and “tone”) of media coverage of research and program activities</li> </ul>	<ul style="list-style-type: none"> <li>▪ CTG invited to become a member of the US NCI clinical trials cooperative group program</li> <li>▪ Unprecedented media coverage of trial results; widely reported press release</li> <li>▪ NSABP B-33 trial is stopped due to the MA 17 trial results</li> </ul>
 3. Engagement / Participation	The characteristics of program participants and clients; number, nature of involvement, and background	<ul style="list-style-type: none"> <li>✓ Level of engagement with other centers, networks, academic institutions, government agencies, etc.</li> <li>✓ Level of engagement by stakeholders / target groups / participants</li> <li>– Level of multidisciplinary and / or multisectorial research activities</li> <li>✓ Level of recruitment and retention of stakeholders / target groups / participants (e.g. junior investigators, researchers, review panelists, etc.)</li> <li>– Level of established external scientific advisory board(s)</li> </ul>	<ul style="list-style-type: none"> <li>▪ A Canada-wide joint press release (CCS, US NCI, PMH) on the trials results was distributed in 2003</li> <li>▪ Higher than projected patient accrual</li> <li>▪ Involvement of international groups like SWOG, ECOG, CALGB, NCCCTG, EORTC, IBCSG</li> </ul>
 2. Activities & Outputs	Implementation data on what the program actually offers	<ul style="list-style-type: none"> <li>✓ Level of research as per internal review guidelines</li> <li>– Extent to which plans, strategies, frameworks, etc. are delivered as per expectations (expected timelines, resource usage and quality levels)</li> <li>– Extent to which governance structure adheres to internal guidelines</li> <li>✓ Extent to which policy and financial decisions are made according to Board / Senior Management / Expert Advisory Committee(s) accepted guidelines and standards</li> <li>✓ Extent to which internal and external communication strategies adhere to internal standards and protocols / policies</li> </ul>	<ul style="list-style-type: none"> <li>▪ Trial unblinding at the time of the first interim analysis as recommended by the CTG’s DSMC</li> <li>▪ Prompt communication of trial results to patients, physicians and the research community</li> <li>▪ Trial funding and management according to CTG standard operating procedures and policies</li> <li>▪ Use of stringent trial protocols (quality assurance) developed by the CTG in collaboration with external experts and partners</li> </ul>
 1. Inputs	Resources expended; number and types of staff involved; time expended	<ul style="list-style-type: none"> <li>✓ Level of human resources (staffing) at all levels (according to norms, vacancies, expectations, benchmarks)</li> <li>✓ Level of financial resources (budgets vs. actuals) at all levels</li> </ul>	<ul style="list-style-type: none"> <li>▪ Level of CTG effort expended</li> <li>▪ \$13 million from Novartis</li> <li>▪ \$5 million per year from CCS</li> </ul>

Figure 6. MA 17 clinical trial outcomes and impacts organized according to the Society’s results chain hierarchy

Figure 6 shows that the MA 17 trial displays evidence of results at all levels of the results hierarchy. This chart validates the initial 'menu' developed in Figure 5.

Research and innovation which preceded (and led) to the MA17 trial along with highlights of its results are traced over time in Figure 7. In figure 7, key activities and results are mapped both vertically, according to the results chain hierarchy levels, and horizontally, across time and the R&D and innovation spectrum. Figure 7 should be read from left to right and from the bottom up. This approach attempts to build in an appreciation of the multi-year build up of research which led to the MA 17 trial, while showing the long-term impacts of the trial. As such, it recognizes the trial as part of a science and innovation pathway of connections. Note that analysts (McDonald *et al*, 2001) suggest downward loop-backs with help support 'upward' progress in such a model. Only one of these loop-backs is shown here, but many others were found.

Figure 7 shows that the Society funded endocrine research and the CTG organized other clinical trials in the time periods leading up to the MA 17 trial. It was not established that the Society-funded research or clinical trials conducted prior to the MA 17 were fundamental to the trial itself. Rather it can be said that through the funding of this research, the Society and the CTG forged important connections in terms of social networks which may have helped to undertake the MA 17 trial in the late 1990s.

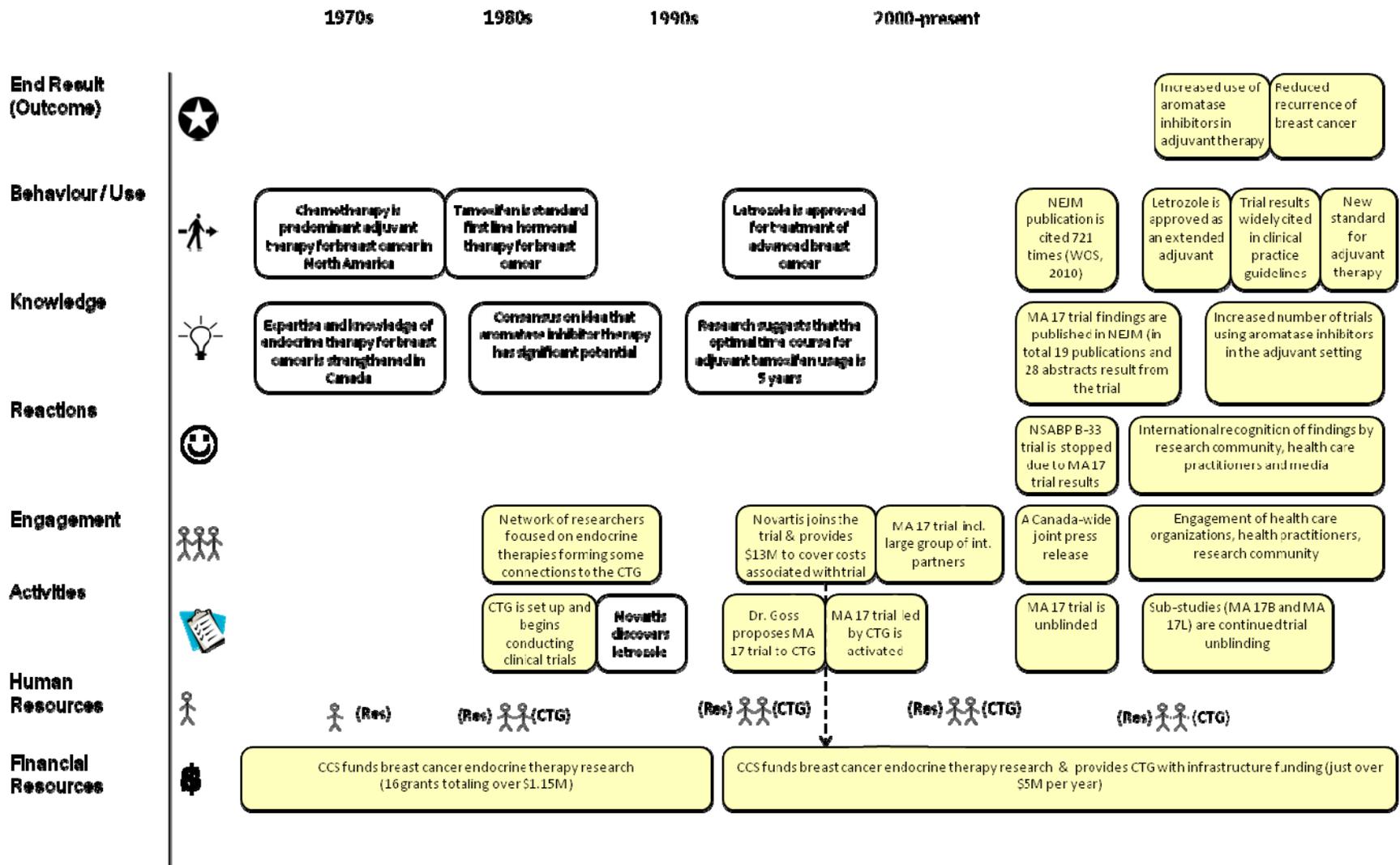


Figure 7. Using the results chain hierarchy to trace the MA 17 trial over time along the R&D and innovation spectrum

Key informants suggest that the MA 17 clinical trial had significant impact on a number of different areas, both directly, in the form of follow-on related clinical trials, and indirectly by encouraging an increased research focus on late adjuvant breast cancer therapy and increasing the use of aromatase inhibitors in all (early and late) adjuvant settings. The results of the MA 17 trial have contributed to changing world-wide standard of care practices that are cited prominently in provincial, national and international guidelines for the adjuvant treatment of breast cancer. In addition, successfully bringing together major oncology groups from North America and Europe increased the CTG's reputation and ability to develop, mobilize and coordinate large international trials.

As shown in this case example, the results chain hierarchy can be used to trace an initiative over time along the R&D and innovation spectrum. Key activities and results are mapped both vertically, according to the results chain hierarchy levels, and horizontally, across time and the R&D and innovation spectrum. This approach attempts to build in an appreciation of the multi-year build up of R&D and innovation while showing both shorter term (more attributable) behavioural outcomes and long-term impacts on mission and (potentially) other elements.

As noted above, plausible attribution can be established by tracking events over time and gathering evidence about connections along a hypothesized change theory. When combined with an established framework such as the Bennett hierarchy of change, this can be a powerful analytical tool.

## **Conclusions**

The measurement of Research Technology & Development and its impact on innovation has gone through phases where high level aggregated metrics gave way to case specific metrics. Both of these applications were used to essentially 'justify' investment. The science management community seems to be poised to move consistently beyond the use of measurement and evaluation for justification to explore how and why innovations occur. This will both help with 'justification,' since impacts will be explained by testable hypothesis (innovation theories) *and* it will help groups to learn about what works for whom in what conditions and circumstances, allowing them to have a greater impact on patients and populations.

For instance, at the Canadian Cancer Society Research Institute, the results of the MA 17 clinical trial case study are being used as a learning tool in a variety of different ways, well beyond those associated with the traditional notion of justifying investment. It has provided insight into the value of funding clinical trials, how to effectively tell research stories and demonstrate impact, and has helped identify gaps in the tracking of the actual take up and use of Society funded research. In regards to the last point, the Canadian Cancer Society Research Institute is now working on ways to better track research impacts, and the results of this case study are providing clues on how best to do this.

A results framework or hierarchy imbued with sound innovation diffusion and knowledge translation theory shows promise as an approach to help transform the evaluation of RT&D and innovation. Some of the key concepts of such an approach include the notions that it should:

- ▶ base the framework on innovation and knowledge translation theory;
- ▶ refer to a hierarchy to sort key indicators;
- ▶ tell a performance story over time; and,
- ▶ include specific context, reach, interactivity, user decision making and behaviours leading to mission and other impacts.

The future may bring an even more results oriented focus. Recent work suggests that an approach which first defines need, then looks at 'market' outcomes linked to *both* an innovation continuum *and* a sector 'value chain' can be helpful. At least one innovation centre is building this kind of thinking into its planning and management with direct implications for evaluation. Work in progress at Forest Product Innovations (FPI) in Canada has drawn on the work of Carlson *et al* (2006) and certain evaluation initiatives to combine the idea of a sector value chain and an innovation continuum (from basic and applied research to development, transfer and commercialization) to plan, measure and evaluate research, technology,

science and innovation investments. Recent work by the National Research Council of Canada is experimenting with a model that starts by asking what behaviors (in different key communities) the Council intends to effect, then proceeds to examine outcomes and impacts.

In summary, RT&D and innovation evaluation has progressed from linear mental models measuring inputs and easily quantifiable research 'products' at a high level, and using broad assumptions to calculate impacts, to approaches which emphasize specific context, reach, interactivity, user decision making and behaviours leading to mission and other impacts.

By drawing more clearly on some fundamentals of evaluation and related theory, RT&D and Innovation evaluation appears ready to experience a break through. Evaluation appears poised to finally go beyond being a mechanism to linearly sum accounts and to retrospectively defend investments from an implied technology 'push' viewpoint. RT&D and innovation evaluation practice has begun to recognize systems, reach and user focussed behaviour changes as keys to both learning and to evaluation. Early evidence suggests that with these changes RT&D evaluation can become an integral part of strategic management.

## References

- Bennett, C 1979. Analyzing impacts of extension programs. Washington, DC: US Department of Agriculture, Science and Education Administration (ESC-57).
- Bennett, Claude and Kay Rockwell 1995. *Targeting outcomes of programs (TOP): An integrated approach to planning and evaluation*. Nebraska, United States: Nebraska Cooperative Extension Service.
- Burstein, H J 2003 Beyond tamoxifen – extending endocrine treatment for early-stage breast cancer. *New England Journal of Medicine*, 349, 1857-1859.
- Carlson, Curtis R and William W Wilmot 2006. *Innovation: The Five Disciplines for Creating What Customers Want*. New York, United States: Crown Publishing Group.
- Davison, Colleen M 2009. Knowledge Translation: Implications for Evaluation. In *New Directions for Evaluation*, eds. J M Ottoson and P Hawe, pp. 75-100. California, United States: Wiley
- Gault, Fred 2006. *Innovation, Indicators and Policy: A Canadian Perspective, The Future of Science, Technology and innovation Policy: Linking Research and Practice*, SPRU 40<sup>th</sup> Anniversary Conference, September 2006
- Gibbons, Michael; Camille Limoges, Helga Nowotny, Simon Schwartzman, Peter Scott, & Martin Trow 1994. *The new production of knowledge: the dynamics of science and research in contemporary societies*. London: Sage. ISBN 0-8039-7794-8.
- Goss, Paul et al 2005. *Randomized Trial of Letrozole Following Tamoxifen as Extended Adjuvant Therapy in Receptor-Positive Breast Cancer: Updated Findings from NCIC CTG MA 17* Journal of the National Cancer Institute, Vol. 97, No. 17, September 7, 2005
- Jordan, Gretchen B 2010. A theory-based logic model for innovation policy and evaluation. *Research Evaluation*, 19(4), October, 263-273.
- Kuhlmann, Stefan and Phillip Shapira 2001. *Learning from Science and Technology Policy Evaluation*, School of Public Policy, Georgia Institute of Technology, Atlanta, USA and the Fraunhofer Institute for Systems and Innovations Research, Karlsruhe, Germany

Mansfield, Edwin, John Rapoport, Anthony Romeo, Samuel Wagner and George Beardsley 1977. Social and Private Rates of Return from Industrial Innovation. *Quarterly Journal of Economics*, 91(2), May 221-240.

McDonald, Bron *et al* 2001, *Introduction to Evaluation - Project Evaluation Training*, Department Primary Industries, Victorian Government, Melbourne Victoria.

McLaughlin, John A and Gretchen Jordan 1999. Logic Models: A Tool for Telling Your Program's Performance Story. *Evaluation and Program Planning*, 22(1), Spring, 65-72.

OECD 2002. *Frascati Manual*. OECD

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

Patton, Michael Quinn 1997. *Utilization-focused Evaluation: the New Century Text*. Thousand Oaks, CA, USA: Sage.

Reed, J and Gretchen Jordan 2007. *Using systems theory and logic models to define integrated outcomes and performance measures in multi-program settings*, *Research Evaluation* 16(3), September.

Ruegg, Rosalie and Irwin Feller 2003. *A Toolkit for Evaluating Public R&D Investment Models, Methods, and Findings from ATP's First Decade*. Maryland, United States: National Institute of Standards and Technology.

Rogers, Everett M 1995. *Diffusion of Innovations*. New York, United States: Free Press.

Rosenborg, Nathan 1983. *Inside the Black Box: Technology and Economics*. Cambridge: Cambridge University Press.

Shodjai, Foad 1996. *Science and Technology Indicators And A Catalog of Major S&T Indicators of Canada*, Centre for Policy Research on Science and Technology (CPROST), Simon Fraser University Vancouver, BC

Teather, George G and Steve Montague 1997. Performance Measurement, Management, and Reporting for S&T Organizations – An Overview. *Journal of Technology Transfer*, 22 (2), June, 5-12.

Wooding, Steven, Stephen Hanney, Martin Buxton and Jonathan Grant 2004. *The Returns from Arthritis Research*. Cambridge, United Kingdom: RAND Europe.

## Notes

---

<sup>1</sup> Publications can often be termed outputs, however if one argues that a publication in a refereed journal constitutes outside acceptance (by definition a change outside of the control of a researcher) then publications may be considered an early outcome.