White paper on

Metrics for the Treatment Sector or Meso Level of the Canadian Health Care System[1]

by

Jerald Hage

Director,

Center for Innovation

University of Maryland

College Park, MD 20742

United States of American

Second draft, 15 June, 2008

1. This second draft has benefited from a number of comments and remarks of the panel.

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

In its sub-title the document entitled Prospectus for a Major Assessment: The Return on Investments in Health Research: Defining the Best Metrics (Canadian Academy of Health Sciences 2007) raises an important question, implicitly if not explicitly: What are the best metrics for measuring the return on investment? This in turn leads to another fundamental question: By what criteria should one evaluate the metrics that are employed? To answer these two questions, I suggest that evaluation metrics for medical research, and beyond this, the evaluation of the returns on investment in scientific and technological research more generally, should have a number of important characteristics. They should be both simple to use and yet fine-grained enough to capture small incremental improvements in health care from specific research studies and at the same time provide the capacity to assess the entire investment in medical research, i.e. be able to function at both microscopic and macroscopic levels. Since there are a number of stakeholders, the range of metrics needs to include not only economic benefits but social ones as well, a point well recognized in the synthesis report of the meetings in Canada to discuss the development of a metric system (Canadian Academy of Health Sciences 2005). The system of metrics should provide a considerable amount of feedback to policy makers and in particular point to how obstacles and blockages that are reducing the effectiveness of investments from medical research can be eliminated. Too often, evaluation studies do not stipulate how returns on investment can be increased, a most important desiratum given the increasing competition for research funds between different and vital sectors of science and technology. And finally, the metric system should reflect the latest advances in evaluation research so that it represents the "state of the art".

The objective of this white paper is to propose such a metric system for evaluating the returns on investment (hereafter ROI) from medical research for the Canadian health care system. The strategic choices of specific metrics in the system are discussed in the first section of this white paper so that evaluators can understand the logic that undergirds the system. The first and most critical choice is the selection of the level of analysis.

Many evaluation systems of scientific research and industrial innovation are constructed at the macro level. Indeed, the focus of the document Developing a CIHR Framework to Measure the Impact of Health Research (Canadian Academy of Health Sciences 2005) and its logic models are largely at this level (see in particularly p. 27). This is a critical level for several reasons, both practical and theoretical. On the practical side, policy makers shape and debate policies at this level. On the theoretical side, the macro level of evaluation fits within the new institutional framework called the national systems of innovation (Nelson, 1993). Recognizing the importance of these objectives, this proposed system of metrics starts with the treatment level within the health care system (see Figure One for definitions of these key concepts) and then suggests how data collected at this level can be aggregated to the national or macro level of policy making, the initial focus of Developing a CIHR Framework to Measure the Impact of Health Research (Canadian Academy of Health Sciences 2005). Therefore, the focus on the treatment sector level allows for retaining all of the advantages of the current logic models and at the same time provides much more information for policy makers. Within health care, the most appropriate term for this level is morbidity sector, which represents the incidence of a particular disease or injury or other kind of health care problem.

Introduction: Strategic Choices and Overview of the Metric System

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

The most important decision is the selection of the level at which data is collected. The argument in this white paper is that the data should be collected at the treatment sector level because it is where the greatest variation occurs. The meso sector level is defined by the *differences* in treatments for various kinds of patients in the health care system, whether for genetic defect (alpha minus one deficiency), injury (post-traumatic stress syndrome), illness (breast cancer), and degenerative process (Alzheimer's disease). An important assumption, but one that is self-evident, is that medical research also varies by the nature of the health care problem and implied treatment in the broadest sense of the term to include prevention and even biomedical and population research that expands knowledge that can be essential for developing

treatment strategies. Since the greatest variation occurs in the nature of medical research for specific kinds of treatments, this becomes the most effective level for feedback to policy makers, who need specific policies rather than one general policy that might be more effective for some areas of medical research and their treatment sectors than others. But its greatest advantage is the simplicity it provides in making attributions that connect the particular research study and its impact on health care and beyond it the various economic benefits that accrue. At the macro level, where various research studies are confounded, even within the same disease category, the correct attribution cannot be made.

Collecting data at the treatment sector level does not mean that various macro level indicators of performance such as contributions to knowledge or broader economic benefits are lost to view. Quite the contrary. These are aggregated across the different treatment sectors. At the same time, this meso sector level of metrics allows policy makers to focus on only one stream of research and evaluate its benefits, scientific, economic and social, if it so desired. One of the intellectual advantages of developing metrics at the treatment sector level is the highlighting of the treatment process consisting of various stages, prevention, diagnosis and prognosis, treatment and post-treatment. Medical research typically impacts on only one of these stages and seldom on the major outcomes desired such as increases in the average duration of life given the morbidity (QALYs). Thus, viewing the treatment as a process allows one to observe the many small incremental steps towards this objective. Therefore, this analytical level allows for a fine-grained system of health care metrics that parallels the treatment process, as is indicated in section two.

The definitions of the different parts of the health care system are listed in Figure One, indicating how complex this system is. The different meso treatment sectors can be aggregated into the national or macro system of health care. The different meso research sectors, one for each of the major treatment sectors, can be aggregated into the national or macro system of medical research. The same could be done for training programs. More critically, although the main focus in this white paper is on the stages of the treatment process, the stages of the medical research process are indicated as well because this

impacts on the quickness with which new treatment protocols are developed and more knowledge about the functioning of the body is accumulated. The former issues are discussed in section two while the latter are the focus of section three.

Figure One

Definitions of Key Concepts

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

Evaluations of scientific research including medical research are caught on the horns of a major dilemma. On the one hand the increasing desires for accountability necessitate quick assessments of pay-off for feedback to policy makers and yet on the other hand the desire for an expanded assessment of the societal benefits requires a longer temporal horizon. Recognizing the difference between potential pay-offs from medical research and the actual benefits that can only occur when the new knowledge is widely diffused throughout the health care delivery system can resolve this dilemma. In evaluation research, this can be called the distinctions between a research discovery or finding, its dissemination, and complete diffusion within the health care system. One of the objectives of this white paper is to indicate how one can measure both the impact of the discovery and the impact of its diffusion. In the second section during the discussion of the metrics of the treatment process this distinction is made, providing a solution to several critical problems in evaluation research: measuring advances in scientific knowledge as distinct from measuring organizational learning, an important frontier issue in evaluation research, and thus representing the "state of the art".

These distinctions also provide some solutions to several intellectual objectives contained in the Canadian Academy of Health Sciences (2005, p. 2). Potential impact of research findings measures the amount of advance in medical knowledge and actual impacts indicate the amount of capacity building. As indicated above, one of the important criteria for the selection of a metric system is: How many intellectual problems does it solve?

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*. This is called prospective evaluation. How this might be accomplished is discussed in the second and third sections of this white paper. With this information, policy makers can do more effective planning (Canadian Academy of Health Sciences, 2007, p. 5)

Another important strategic reason for focusing on the meso level of the treatment sector and then aggregating to the macro level is the flexibility it provides not only policy makers but also evaluators. Above, we described the flexibility of temporal horizon in conducting the evaluation; it can handle both short- and long-term assessments. But this is not the only kind of flexibility. Another is the specificity of the focus. Since the treatment sector is defined by the similarity of treatments or technologies and the homogeneity of clients or customers, one can select different levels of homogeneity for evaluating ROI from medical research. One can make finer and finer distinctions between the similarity of treatments and patients. One might choose to focus on bullus emphysema at one level or another level, one of its causes, alpha minus one deficiency, a sub-division of the causes and thus an increase in the homogeneity of the patient pool. This distinction becomes important when the research is in fact concentrating on only one of the potential causes while developing an effective treatment, as is frequently the case.

Perhaps the most important strategic advantage of developing health care and economic indicators at the treatment level is the simplicity and ease of coding the results of specific research studies. At this level, it becomes much easier to make attributions or linkages between investments in a particular kind of research and its potential, or actual pay-off, for both health benefits (better diagnoses), economic benefits (reductions in treatment costs) or societal benefits (reduction in working days lost) (Executive Summary, pp 2- 3 Canadian Academy of Health Sciences 2005). This document (p. 8) expresses concerns about tracing the linkages between research outputs and health care impacts, especially if the knowledge develops incrementally over an extended time period. Focusing on the meso level or the treatment sector of the health care system solves this problem.

This tight link between medical research and health care impacts in the treatment process is not the only cognitive advantage of a meso system of metrics for evaluating ROI. Another is that it calls attention to how the medical research for a specific treatment is organized. Borrowing ideas from the well-developed literature on how knowledge evolves in the scientific-industrial innovation literature (Hage and Hollingsworth, 2000; van Waarden and Oosterwijk, 2006) provides insights about the

processes of differentiation in medical research organizations, in particular dedicated institutes, that may be leading to gaps that are retarding the development of radically new treatments. If so, this becomes a critical kind of feedback to policy makers.

Strategically, the treatment sector level lies between the macro level of the policy makers and the micro level where research is conducted and patients are treated. It is the missing link that connects these two levels and without research at this level, one can not easily select the correct health care policies or understand what might be various kinds of blockages and obstacles to the creation of new treatments.

This connects to the final criterion for the evaluation of a system of metrics based on the differences between treatment sectors and that is whether it represents the "state of the art". Two new thrusts in evaluation have been advocated by Europeans (Arnold, 2004; 2006). The first thrust is the importance of identifying obstacles and blockages that prevent or retard the rapid development of new innovations, that is treatments and in particular radically new ones. As indicated in section three of this white paper, these obstacles and blockages are the probable causes of the gaps in the rapid development of new treatments. Although their focus is on the evaluation of science and technology, the same logic can be applied to how treatments are organized and whether or not the different stages in the treatment process are well connected. Some of the metrics suggested in Figure Two in the second section indicate how this might be assessed.

The second thrust is the importance of allowing theory to inform the evaluation and hopefully that evaluations of science and technology including medical research start contributing to the construction of theory. While evaluations of medical research have their own specificities, one reason this white paper has stressed the importance of studying meso sector or treatment level, it is also true that different theoretical formulations such as the idea innovation network theory (Hage and Hollingsworth, 2000) can provide cognitive maps that raise important questions and provide new insights. These are the advantages of this proposed system of meso level metrics for determining the ROI from medical research. In other words, new developments in social science theory can inform the evaluation and the evaluation can help contribute to the development of new social science theory. We already have some suggestions of this in

the ways in which advances in knowledge and organizational learning can be assessed with this system of metrics.

Overview of the Metric System

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

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

The issue is how do these different dimensions related to each other. The discussion of these metrics does not follow the logic models provided in Canadian Academy of Health Sciences (2005, p. 27) but this is done for a specific reason. A typical logic model would probably begin with the metrics of research investment since these provide the resources for the creation of the health care impacts and then move to the metrics of health care impacts. I have reversed this logic because the major issue is what are the indicators of health care impacts and in particular what are the advantages of a fined-grained approach at the treatment or meso sector level of the health care system. It seems to me that this is the most important part of this exercise, demonstrating the ROI of medical research. It depends upon the number and variety of health care impact indicators that one has developed.

Throughout the discussions of these metrics, a continual concern is to provide a number of policy feedbacks so that the performance of the health care system can be improved and ROI in medical research can be increased. Indeed, as I have indicated, one of the advantages of studying the treatment or meso level within the health care system is the many useful feedbacks to policy makers including whether some components of the

treatment process are ignored, some arenas of medical research receive little investment and identifying gaps in the production of radical new treatments in the networks connecting different kinds of medical research.

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

The elaborateness of this list of this of metrics helps speak to the quite varied concerns of different state holders, and surprisingly simplifies obtaining answers to many of the questions posed in Canadian Academy of Health Sciences (2005, p. 1), which are consistent with those of Buxton and colleagues (1994) and as modified in the current Canadian Academy of Health Sciences (n.d.): advancing knowledge, informing decision-making, health impacts, and economic impacts.

What this white paper does not accomplish is also important to state at the beginning. The paper is focused on developing metrics, not the methods that one would use to implement some evaluation using these metrics (see Buxton, 2007; Hanney, 2007; Wooding, 2007 for various examples of methods). This would require another white paper, particularly as there are a number of alternative research designs that could be used to collect the necessary data. In addition, the Canadian Academy of Health Sciences (2005: p. 32) has an extended list (also see Hanney et al., 2004). At various points, suggestions about methods are made but it is not considered to be part of what this white paper is intended to accomplish. Nor does this white paper consider the issue of how to evaluate the relative priorities of medical research broadly conceived in comparison to research in the physical sciences, the military or any of the other national goals of the

Canadian government. However, the logic of the process used in this white paper can be applied to other national goals, their delivery systems and research investments, to achieve these goals. If this were done, then one could compare the relative ROIs in different national sectors of concern to the government. Clearly, this is considerably beyond the scope of this present exercise.

Metrics of Health Care Impact

Economists have developed a quite elaborate classification system of industrial product sectors including market niches within them. But the same effort has not yet been applied to the non-economic service sectors such as health, education, and welfare to say little about the new national missions that need to be distinguished such as national security, global warming, etc. Despite this, the same logic for distinguishing between product markets can be applied to the classification of service sectors. It is the differences between the kinds of patients and the kinds of treatments including technologies such as machines, procedures and naturally the human expertise that allows one to observe distinct treatment or morbidity sectors. Some might question the comparisons between the treatment sectors in the health care system and the industrial sectors within the economy. But by thinking in terms of analogues, cone can develop a number of insights. Admittedly, the insights most always be carefully adapted to the specific circumstances to observe some of the more striking differences as well.

One might ask why distinguish separate sectors within the health care system? Since many different treatments are housed in the hospital, the concept of distinct treatment or morbidity sectors may appear to be strange to health care professionals. But it is precisely because of this fact that one needs the idea of distinct treatment sectors so that one can more easily establish a linkage between a specific research finding or body of research and its health care impact even if much of the diagnostic equipment resides in the same place as other treatment sectors. Indeed, this is one of the more interesting planning issues, especially given the problem of intake, when it may be better to provide specialized clinics for particular stages in the treatment sector. Recognizing the alternative treatment systems allows for comparisons across the research findings of the 13 major Canadian Research Institutes, some of which reflect particular kinds of populations (aboriginal, elderly, child, gender), some of which reflect specific systems

within the body (circulatory, neurology, musculoskeletal) and others specific arenas of research (basic such as genetics or service provision as in the health services and policy institute) and despite the name of the specific institute might be involved in research relevant to one or another of the other institutes.

But this is not the only reason. Highlighting the treatment sector or the specific morbidity calls attention to parts of the treatment process that may require strengthening via research on service delivery in the Institute of Health Services and Policy Research. At minimum four components or stages in the treatment process can be discerned: the prevention stage; the intake and assessment stage including diagnosis and prognosis; the intervention stage including hospitalization; and the post-intervention stage including rehabilitation and long term care when appropriate. Prevention is placed prior in time with the simple assumption that if prevention can be successful, then the treatment process is unnecessary.

Metrics for the Four Stages: prevention, intake and assessment, treatment and post-treatment including long term care

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

The four stages used to describe the treatment process are the same as those in the Canadian Academy of Health Sciences (2005, p.). Each of these four stages

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

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

The treatment process logically begins with the stage of prevention. In this stage, I suggest two metrics for measuring health care impact. The first and most obvious one is the relative effectiveness of prevention. Prevention protocols that stopped smoking and discouraged teenagers from starting smoking have steadily improved over the years. As more and more people adopt healthier eating habits and exercise, one observes the decline in the severity of various health problems associated with aging. Perhaps the most dramatic examples of prevention are the development of vaccines that eliminate a specific morbidity such as small pox, polio, and, we hope, some day AIDS.

Intake and assessment is a particularly interesting stage. Within this category, again, I am suggesting three metrics for measuring health impacts. The speed of the diagnosis can be strongly impacted upon by either the adoption of highly specific screening techniques and/or the rearrangement of the delivery system of health care so

that there are more points of contact at which a quick diagnosis can be made.

Figure Two

Metrics of Health Care Impacts

<u>Stages of the Treatment Process</u>

Prevention

- **Percent increase in the effectiveness of prevention intervention (decline in incidence of morbidity)**
- **Percent decline in severity of incidence of morbidity**

Intake and Assessments

- **Percent increase in the speed of diagnosis (reduction in the number of tests and their duration)**
- **Percent increase in the accuracy of diagnosis (reduction in false positives or negatives)**
- **Percent increase in accuracy of prognosis (duration and quality of life, etc.)**

Treatment Interventions

- **Percent increase in the speed of treatment intervention (wait time in the emergency room)**
- **Percent decrease in the length of treatment**
- **Percent decrease in side-effects of intervention and/or their severity**
- **Percent decrease in opportunistic infections during treatment intervention**
- **Percent increase in the quality of life during treatment (reduction in invasive procedures, increased opportunities to be treated as an outpatient, reduction in pain during major interventions, etc.)**
- **Percent increase in success rate of intervention**

Post-Treatment Interventions (rehabilitation and long term care)

- **Percent increase in the speed of the rehabilitation intervention**
- **Percent decrease in length of rehabilitation and long-term care**
- **Percent increase in the quality of life during rehabilitation and after care (decrease in the pain of rehabilitation procedures, reduction in invasive procedures, opportunities to be treated as an outpatient).**
- **Percent increase in success rate of rehabilitation (DALYs) or increase in physical (vision, hearing, thinking, movement, dexterity) and psychological functioning (cognitive processing, speech, memory) after stroke or injury that impaired functioning**

Summary Output Measures of the Morbidity Sector

- **Percent increase in the average duration of life given the morbidity (QALYs)**
- **Percent increase in the quality of life after interventions (reduction in recurrences, continuity in mobility, reduction in constraints of life style, etc.)**

Knowledge about the Health Care Problem

- **Percent change in the understanding of the causes of the health care problem**
- **Percent increase in the sub-categories of the health care problem**
- **Percent increase in the understanding of the relevant biological and psychological processes of the body relevant to the health care problem**

The importance of eliminating false negatives and false positives in the diagnosis is well understood as is the importance of the accuracy of the prognosis._Accurate prognosis can lead to quite different decisions about whether it is worth starting a treatment. Many patients will accept a death sentence provided there is good control of pain. Much research has cast doubts about the advantages of screening precisely because of the inaccuracies of the diagnosis, too many false positives and false negatives (e.g. Marshall. 2008 on mass screening of high risk lung cancer patients). But gradually through decision analysis such as an example Goldie's work, considerable progress is being made in how to effectively screen for certain cancers.

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

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

One can use two summary measures of the four stages: the average increase in the age-sex standardarized duration of life for the specific morbidity and an increase in the quality of life after intervention. The former measure is usually only useful when evaluating a stream of research because most research studies do not impact on this

except only incrementally if at all. The second summary measure, the quality of life after intervention, is discussed below in a special sub-section.

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

Research projects in the disparate institutes can concentrate on one or another stage of the treatment process and the distinctions by stages allows us to recognize their contributions to the overall health of the Canadian population. For example, the improvement in the duration of life of individuals with cardiovascular problems is a consequence of three distinct kinds of changes: faster and better medical interventions after a heart attack, drugs to reduce some of the causes such as cholesterol and blood pressure, and behavior changes involving diet and/or exercise (Canadian Academy of Health Sciences, 2005, p. 13). Each of these different contributions is captured in this system of health care impact metrics.

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

At the meso sector level, these metrics are highly flexible and can be used with micro time, say three to five years or with macro time, three to five decades. In the former instances one examines a few research studies and in the latter one can also

evaluate how a whole program of research, e.g. cardiovascular disease treatments that has unfolded over three or more decades (Canadian Academy of Health Sciences, 2005, pp. 10-1). As was suggested at the beginning of this white paper, policy makers in certain circumstances would like immediate feedback relative to options as to how money should be invested. This fined-grained approach allows for this.

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

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

Special Issue: knowledge about the health care problem

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

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

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

Perhaps the greatest demonstration of these two evolutionary processes towards more complexity and the need for more customization in treatments is reflected in the stream of research about the biological functioning of the body, which provides the background for doing clinical research and may in the long-term lead to more effective interventions. The clearest example of this is the double helix description of DNA and

the demonstration of RNA followed by the decoding of the genome, which in turn has led to a renewed recognition of the complexities involved. Molecular biology has opened the door to the development of gene therapies and the movement towards the customization of treatments. Some recent examples of the kinds of general research in the biology of the body reported in Science recently includes research on the self-organization of proteins (Lutkenhaus, .2008) and the enigmas of blood clot elasticity (Weisel, 2008).

Special Issue: measuring the quality of life

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

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

What are the behavioral indicators that can successfully model the perception of a high quality of life? Among others for measuring the quality of life during treatment are reductions in invasive procedures, opportunities to be treated as an outpatient, reduction in pain during and after major interventions, etc. One example of the first indicator is the substitution of laser surgery for the eyes rather than the more invasive forms of surgery. The use of drugs to prevent surgery has recently been discussed in the case of prostate cancer (Kolata, 2008). But in this instance the potential long-term side effects are unknown and unlikely to be studied. The movement of treatment out of the hospital into the outpatient clinic also has the same consequence. Most people find hospitals are scary

places to say nothing about their inherent dangers associated with opportunistic infections. Finally, the successful management of pain is one critical issue with most patients. When it is reduced without side effects such as dependency, then most patients would consider this as an improvement in the quality of life. Speed of treatment and speed of post-treatment rehabilitation probably also can contribute to the perception of improved quality of life since both waiting, including waiting for the diagnosis and prognosis are stressful times for the patient and his/her family.

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

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

A research project might have implications for more than one aspect of the treatment process, e.g. a new experimental treatment for melanoma at the National Institute of Cancer in the United States considerably reduces the amount of time spent in the hospital, increases the success rate from 15 percent to 50 percent, and improves the quality of life during the treatment. Under these circumstances, one adds the percent

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

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

Potential vs. Actual Benefits

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

focuses on how best to diffuse advances in medical knowledge so that potential benefits are actually realized.

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

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

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

The distinction between potential and actual benefit speaks to two quite different policy objectives in measuring ROI. The potential impact is a measure of the amount of

scientific learning or the extent of the knowledge advance that has occurred from a specific research study or even a whole research program. The actual impact is a measure not only of the extent of diffusion of knowledge but reflects capacity building in the health care system. As a higher percentage of the appropriate health care personnel have learned the new protocols, then the greater the advance in the human capital for that specific morbidity treatment system. Another advantage of clearly separating these two kinds of indications of "pay-back" is that they address some interesting problems in the sociology of science and organizational literatures. *Advances in treatment knowledge represent a measure of scientific learning while diffusion of treatment knowledge represents a measure of organizational learning.* Both of these measures are different from those listed in Figure Two, which focus on the knowledge background of a specific morbidity.

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

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

macro assessment even if it is based on meso treatment sector metrics. Given this concern, the next topic is how to aggregate from the meso to the macro.

Aggregating Across Treatment or Morbidity Sectors and Policy Feedbacks

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

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

At the same time, these percent *changes* need to be weighted by the caseload or relevant population in the morbidity sector, especially when computing the economic and social benefits of investments in medical research. Genetic or birth defects involve small numbers whereas cancer patients. especially in the more common cancers such as lung or breast, involve thousands. An alternative procedure is to assign weights according to the priorities of the government or that reflect the dominant values of the Canadian society (Canadian Academy of Health Sciences, 2005, p. 28). Space does not allow me to explore a variety of various kinds of weighting systems that recognize, for example, the greater difficulty in solving some problems such as the degenerative processes associated

with aging, but regardless of which system is used, the key point is that weighting allows one to build in specific kinds of values about the importance of particular streams of research, another way in which policy makers can be informed.

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

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

The treatment system and the research system or knowledge production system are quite distinctive systems. They should be analyzed separately so that linkages between them can be better understood, our next topic.

What are the Components of the Knowledge Production System?

Metrics for Investments and Network Gaps

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

One of the most important reasons to measure how medical research is organized is to provide feedback to policy makers on how the research system can be made more effective. In particular, there has been concern expressed about how to enhance capacity to innovative (Canadian Academy of Health Sciences, 2005, p. 8). This can only be accomplished if there is a clearly image of how innovations are produced in a knowledge production system or what Hage and Hollingsworth (2000) call the idea innovation network. While there has been some discussion about the utilization of research in policy making including how to measure this kind of impact (Hanney, 2007) and certainly influencing policy makers is one of the crucial elements in the pay-back model, there is little discussion *about what kinds of information should be provided to policy makers*. In other words, what do policy makers need to know? It is my contention that they need at minimum to know three things: (1) expenditures by morbidity, (2) expenditures relative

to specific objectives defined by the metrics in Figure Two, and (3) knowing how to better organize the research system relative to these metrics. The organization of medical research is as important as how to better organize the treatment system, the objective of the Institute of Health Services and Policy. And indeed, the absence of research within this institute relative to certain metrics may be critical for improving the speed of treatment and post-treatment interventions. Therefore, I propose a system of metrics for monitoring investments in different kinds of knowledge production defined by morbidity and metric objective within and the detection of problems in the organization of medical research that results in it being less effective than it could be for achieving the objectives outlined in Figure Two.

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

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

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

Research then provides advances in knowledge that in various ways improve the capacity to make diagnoses, treatments, and provide better follow-up care. Within this general category of advances in medical knowledge, there are distinctive kinds of research, some of which we have already touched upon in our discussion above of the nature of the research institutes. Just as we needed a fine-grained set of health care impact indicators, we also want an elaborated set of research arenas to capture the distinctive contributions of the various Canadian institutes. One such scheme is provided in the

Hage and Hollingsworth's (2000) idea innovation network theory of radical innovation, which modified and built upon the original insights of Klein and Rosenberg (1986). The scheme identifies six ways in which knowledge advances.

Some health care policy makers might question the use of a theory developed by measuring the relationship between scientific knowledge and industrial innovation but the advantages of it are that it allows one to think in new ways about how medical research is organized and in particular whether enough attention is being devoted to specific morbidities and most critically specific metrics within them. Transposing the names used in the industrial innovation literature to terms that are more appropriate for studying ROI of medical research, the six arenas are basic scientific research, applied or clinical research, treatment protocol development, health care system research, quality of care research, and "commercialization" including diffusion of findings research. Figure Three lists three separate metrics or indicators, the first the amount of funds invested in this specific research arena, the second the number of personnel who are engaged in this research, and the third the number of researchers engaged in international teams. Both the second and third are measures of the training of personnel via participation in research and could easily include not only post-docs but also medical students and even undergraduates. One of the concerns of the document Canadian Academy of Health Sciences (n.d.) was the building of capacity. This is the second measure of this capacity building, since the diffusion of treatment knowledge also impacts on this aspect of the health care delivery system. But rather than the building of capacity of health care personnel, which is what the diffusion of treatment knowledge measures, here our concern is the creation of greater research skills. The assumption is that the best training occurs while actually conducting research. If one wanted to, one could also count the number of individuals enrolled in specific training programs for particular morbidities, but this data is probably not readily available.

The four pillars of Canadian research--biomedical, clinical, health services and policy research--that cut across the various institutes do not exactly map on these six arenas but there is still a considerable overlap nevertheless (Canadian Academy of Health Sciences, 2005, pp 21-22). Biomedical research most closely fits basic scientific research

as defined in Ibid (p. 21) and includes not only studies of the eight systems of the body but also the etiology of disease and the fundamental aging processes. Likewise population research such as epidemiological studies can focus on this. One of the more interesting new areas is research that contextualizes treatments on the basis of genetic predispositions or other factors in the body. As we have observed, the movement of medical research and indeed a measure of the knowledge base is how many distinctions can be make; customization of treatments represents the extreme in this. Clinical research is also defined in Ibid (p. 21) broadly and fits the applied research category. Health services and policy research is where issues about the organization of the treatment system and quality of care can be examined. Both the metrics of speed and quality can be improved via research on health services and how they should be distributed in Canada and what are the best approaches with particular categories of individuals. Finally, population research and public health research can also focus on problems of how best to diffuse knowledge advances given the differences in population associated with social class, gender, ethnicity, and education as well as developing more effective prevention treatments.

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

Why develop metrics of investments in the six arenas for each morbidity or treatment sector and beyond this metric outcome? Besides the human capital implications contained in the second and third indicators within each of these arenas, there are a number of reasons as to why metrics of investments have to be part of the conceptual accounting scheme. *First*, to return to the implications of these metrics for policy makers, the fine-grained approach exposes gaps in funding. This is especially true when one shifts to the ignored level of the specific metrics within the morbidity that one would like to have as objectives. For certain morbidities there may be a total absence of

research on the protocol development or how best to diffuse the knowledge that is gained. A major desiratum of the Canadian Academy of Health Sciences is to develop skilled researchers. This fine-grained list of six kinds of researchers also highlights lacunae that may be affecting the performance of the health care system in adverse ways.

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

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

Third, the third metric within each arena is measuring the extent of international cooperation or participation in Canadian research within an arena of a specific morbidity. International teams provide a number of advantages that are well-known: upgrading of the human capital of the researchers, spreading scarce resources so that more can be learned with the same amount of investment, and depending, making more substantial contributions to medical knowledge. The value of these collaborations is stressed in Canadian Academy of Health Sciences, 2005 (p. 20). There value is enhanced when one focuses on the particular metric outcome of a specific morbidity because it is so difficult for the Canadian medical research system to examine all metric outcomes in all morbidities. Indeed, as the complexity of medical research increases and in particular the movement towards customization, cooperation across national borders becomes more essential. The fined-grained approach of the health care metrics allows policy makers and medical researchers in Canada to select their partners with a sharper focus. For

example, given the new one billion dollar center in Germany consecrated to studying dementia, Canada might prefer having a relationship with that country for that specific morbidity.

Figure Three

Metrics of Medical Research Investments

Relative to a Specific Morbidity Sector and/or Metric Outcome

Basic scientific research
- **Canadian dollars invested**
- **Number of personnel performing the research**
- **Number of personnel participating in international teams**

Applied or clinical research
- **Canadian dollars invested**
- **Number of personnel performing the research**
- **Number of personnel participating in international teams**

Treatment protocol development (tools, machines, techniques, procedures, etc.)
- **Canadian dollars invested**
- **Number of personnel performing the research**
- **Number of personnel participating in international teams**

Health care system research including manufacturing of tools and equipment
- **Canadian dollars invested**
- **Number of personnel performing the research**
- **Number of personnel participating in international teams**

Quality of health care research
- **Canadian dollars invested**
- **Number of personnel performing the research**
- **Number of personnel participating in international teams**

Commercialization and diffusion of knowledge research
- **Canadian dollars invested**
- **Number of personnel performing the research**
- **Number of personnel participating in international teams**

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

most critically teaming allows for more radical breakthroughs in either scientific contribution. Knowledge of the morbidity, protocol development or commercialized medical product, assuming that property rights can be distributed appropriately.

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

Special Issues: detecting gaps in medical research by health care impact and arena of research

As has already been suggested, two important objectives of any metric system for evaluating ROI from medical research is that it provide useful information to policy makers and that it represent the "state of the art". Both of these objectives are achieved when the system of metrics allows one to detect gaps in medical research. The first issue for policy researchers to ask is are there any investments within a morbidity for each of the health care impacts listed in Figure Two? Obviously, I do not have detailed knowledge of the investment portfolio of the Canadian medical research system and therefore cannot point to any particular examples of lacunae. The key point is that the fine-grained list of twenty health care metrics should allow for a careful evaluation of funding for each morbidity. In any case, probably medical researchers when applying for funds report the expected impacts on the suggested list of health care impacts.

The second question to ask is whether for a specific metric outcome, the six arenas are represented in the research team. The four pillars do raise some questions as to whether enough attention is being given to developing the necessary skills for the creation of medical machines, drugs, surgical tools, disability enhancing tools that can be commercialized, a concern of the Canadian Academy of Health Sciences (2005, 17). The potential role of drug treatments, if they could be developed, to replace surgical treatments for specific morbidities would represent a large market. Canadian Academy

of Health Sciences classifies commercialization of research as part of clinical research. In contrast, the idea innovation network theory argues that this is a separate set of research skills that should be developed. Protocol development, especially for innovative products that can be commercialized, may require the participation of medical engineers, who are not normally included in the clinical category of researchers. Besides the development of new drugs, there are many other kinds of products such as hospital equipment, measurement instruments, surgical tools, etc. For example, in the rehabilitation of accidents, and especially injuries of American solders in the Iraq War, there are opportunities for Canadian medicine if they desire to develop the special kinds of mobility aids to walking, talking, seeing, hearing, etc. that are needed. Many of these aspects of new treatment protocols can be commercialized, although the best methods for doing so are the topic of research in the sixth and last arena. This research arena of protocol is the one most directly concerned with the problem of innovation.

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

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

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

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

Metrics for Knowledge Contributions

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

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

in treatment knowledge relative to a specific morbidity a more useful feedback for policy makers because the time lags are less.

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

Figure Four
Metrics for Knowledge Contributions

Metrics for a morbidity sector	*Metrics for international recognition*
1. Number of publications	**1. Number of international publications**
2. Number of citations	**2. Number of international citations**
3. Number of patents	**3. Trade balances for patents**
4. Number of patent citations	**4. Number of international citations**

Although prizes are an example *par excellence* of international recognition, this metric is not included. The problem with this specific indicator is the very long lag time between the completion of research and its recognition by one of the major biomedical prizes (Horowitz, Lasker, Nobel Prize in Medicine and Physiology or in Chemistry, when relevant, etc.), which is usually anywhere from 15 to 30 years.

The Metrics of Network Gaps

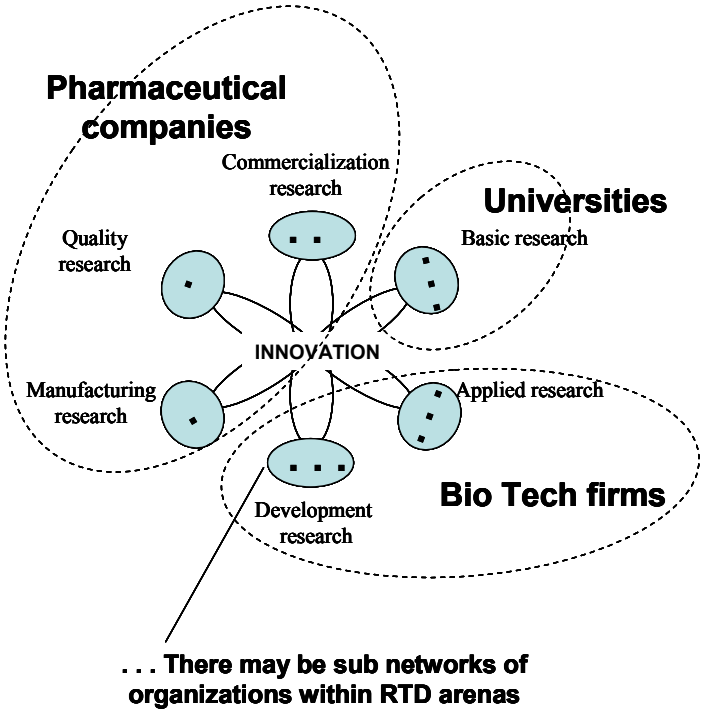
Above, I have suggested that one of the most important kinds of feedbacks for policy makers is to know how to better organize the research system. In particular, it is

critical to identify gaps in funding, and even more importantly organizational and network obstacles to doing good research, as Arnold (2004) has argued. As has been suggested the metric system should represent the "state of the art". The importance of investment funds and having capabilities in all six arenas of the idea innovation network theory, especially relative to the specific metrics listed in Figure Two for each morbidity affect the amount of ROI and the speed with which new and in particular effective protocols are developed. This is the basic argument of the idea innovation network theory (Hage and Hollingsworth, 2000); for an example of the thinking in biotechnology and pharmaceuticals see Figure Five. If any of the links between the research organizations that handle particular arenas of research are broken or those arenas of research are not represented, the radical innovation becomes quite difficult.

Figure Five

Idea Innovation Network Theory (Hage and Hollingsworth, 2000):

Bio-tech/Pharmaceutical Sector



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

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

The idea innovation network theory makes a number of predictions about how the idea innovation network evolves with the growth in knowledge. Specifically, it argues that one needs more complex research teams to increase the speed of radical advances in the various outcomes listed in Figure Two. Creating networks that link the arenas together tightly allows for a higher rate of major breakthroughs in the development of medical protocols, including their rapid diffusion throughout the health care system. Admittedly, the objective of research is not always a major breakthrough or radical

innovation because most treatments advance incrementally and frequently through experience rather than research. But it is also true that, when health care crises such as pan-epidemics occur, AIDS being one them, radical innovations and speed in the development of treatment protocols become critical.

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

The combination of all six arenas for high priority research is the major thrust of the idea innovation network theory. The most effective way of combining them is in complex research teams that include researchers from all arenas. At the same time, their integration into this quite complex research team is not self-evident. Hence the concern about not only gaps in the connections between these different arenas of research but also in the integration of the complex research teams that connect research arenas. Furthermore, the quality of the integration in the networks that connect these arenas affects the speed with which new ideas are developed. For government policy makers

then this idea innovation theory allows them to look for communication problems within research teams, another kind of gap listed in Figure Six.

Figure Six

Metrics for Detecting Network Gaps

Metrics for detecting network gaps between research arenas

1. **Absence of linkages between basic and the other five**
2. **Absence of linkages between applied and remaining four**
3. **Absence of linkages between protocol development and the remaining three**
4. **Absence of linkages between service provision and the remaining two**
5. **Absence of linkages between quality of care and "commercialization"**

Metrics for detecting weak linkages or communication gaps

1. **Absence of complex research teams**
2. **Low communication rates within complex research teams**

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

What are Returns on Investment?

Economic and Societal Benefits

One of the advantages of having a fine-grained list of health care impacts is that it suggests a number of ways in which economic gains can be computed. The cost benefit studies cited in the Canadian Academy of Health Services (2005, pp. 10-11) such as cardio-vascular and diabetes evaluations usually cover two to five decades. However,

policy decision-makers also need something much more immediate in evaluations of ROI. The advantage of having health care impacts in the treatment process by stages as well as increases in the knowledge about the morbidity is that it not only allows one to draw more clear linkages between specific research studies but it allows one to have much more immediate assessments.

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

Economic Returns

The metrics of Figure Two in most cases suggest the economic gains, which are listed in Figure Seven. As in Figure Two, the indicators are listed under each of the four stages. But the economic benefits tend to become greater the higher on the list that a health care metric has been impacted. Also, there are fewer economic metrics than health care impacts (16 vs. 20) because several health care impacts can result in the same economic benefit and since the economic benefits of additions to knowledge about the health care problem are difficult to estimate without being able to trace directly from the knowledge to a specific health care impact. This problem is exemplified when researchers announce that some day their particular finding might result in a new vaccine or gene therapy.

Some discussion is necessary along with some examples. Developing vaccines for certain morbidities has had profound impacts on the cost of treatments, the saving of lives, and the quality of life and thus there is a multiplier effect in the economic benefits. The potential economic gains via prevention are quite large even if the actual benefits in this stage of the treatment process are small when prevention requires that individuals change their behavior. For example, a decrease in the incidence of AIDS means a considerable number of illness days saved for each individual. The value can be

computed from clinical records of the average yearly cost of treatment even if this is largely a regime of drugs. The same is true for the reductions in the severity of the morbidity. A little exercise reduces frequently the severity of the heart attack, which in turn has a number of economic benefits. As we have seen, recent research indicates that exercise combined with a proper regime of food actually changes which genes are operative and which are not.

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

The next set of metrics for health care impacts on the treatment process provides a variety of ways in which economic gains can be computed. The value of the days, months, years added to life because of successful treatment interventions or QALY's is probably the core element of how medical research can be assessed economically. The speed of intervention can influence the likelihood of a successful outcome. Even if not, the speed reduces costs for the patient's family as they wait. The issue is how should the value of years added to life be computed? On the basis of average salary or in terms of some stipulated value of human life? Different evaluators can make different judgments but, regardless of the decision, the logic remains the same: the economic value of the time saved. Reductions in treatment costs because of shorter hospital stays, fewer side

effects and fewer opportunistic infections are relatively straightforward and would be based on the average cost of a hospital stay per day or visit to a clinic. However, when drugs are substituted for surgery, there are issues about the costs of any side-effects from the drugs.

Figure Seven

Metrics of Economic Benefits from Health Care Impacts

Stage of the Treatment Process

Prevention

- **Value of illness days saved from decline in morbidity incidence**
- **Value of reduction in cost of treatments for less severe morbidity incidence**

Intake and Assessments

- **Reduction in the costs of tests for diagnosis**
- **Reduction in costs of false positives or negatives**
- **Reduction in the costs of futile interventions**

Treatment Interventions

- **Reduction in the patient's costs of waiting**
- **Value of life days added by successful interventions**
- **Value of reduction in treatment costs because of reduction in length of treatment (e.g. hospital days)**
- **Percent decrease in treatment costs of side-effects of intervention and/or their severity**
- **Percent decrease in costs of opportunistic infections during treatment intervention**
- **Percent decrease in treatment costs because of less invasive procedures, shift from hospital to outpatient**

Post-Treatment Interventions (rehabilitation and long term care)

- **Value of days saved in rehabilitation and after care**
- **Percent decrease in treatment costs because of less invasive procedures, shift from rehabilitation hospital to outpatient care**
- **Value of increased mobility of all kinds after rehabilitation**

Summary Output Measures of the Morbidity Sector

- **Value of increase in the average duration of life given the morbidity**
- **Value of absence of reoccurrence in health care costs and increase in the quality of life after interventions**

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

hospital and most importantly in long-term care represent real and critical economic gains.

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

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

Finally, the overall metrics, such as average increase in life expectancy relative to specific morbidity, are usually computed in cost-benefit studies of medical research such as those described in the Canadian Academy of Health Sciences (2005, 10-13). The second metric of improved quality of life is more difficult to assess. As we have seen, some of the behavioral indicators of quality can be quantify as economic gains but not all. The costs of living with constraints, for example, is a difficult one to assess. There are methods for doing so but these are beyond the scope of this paper. In any case, the other metrics are being emphasized as way of not having to make this assessment with surveys,

one easy way in which it can be done, and instead on the basis of changes in the other metrics.

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

Societal Benefits

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

For the purposes of this exercise, I have deconstructed society into four distinctive realms: education and scientific, political, economic, and health and welfare. There are other classification systems including my own of eight (Hage, 1972) but these four reflect those that are most critical for the purposes of government policy. These institutional realms each have performance outputs. I have only suggested two but, I think, the most important two in each institutional realm. In the educational and scientific realm, one major performance output is the upgrading of human capital of both professionals and researchers. One could construct an index of the number of new protocols learned and the number of new research projects participated in. The second output reflects the recognition of Canadian medical science. Again, one could construct an index of increased international recognition of Canadian medical science based on citations to papers and patents and number of foreigners that participate in research teams. The weights involved in the construction of these indices are again beyond the scope of this paper. I am merely suggesting some ways in which one could proceed.

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

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

Some may be somewhat surprised that I have included the health and welfare realm since basically the whole white paper has been on a variety of metrics involving health status. What else is there left to measure? The answer is the increased equality of health care among social classes and gender groups as well as meaningful distinctions with Canadian society. Research on the diffusion of knowledge throughout the treatment system including how best to reach various groups with poor health is generally a priority

of governments. This is frequently the kind of contribution that the pillar of population research can make. Although national health care systems are usually established to create equality of health care, they frequently do not, as the exhaustive study of equality in the British health care system indicates (Hollingsworth, Hage, and Hanneman, 1989). One way in which the decrease in inequality could be measured is by the reduction in the differences in duration of life.

Figure Eight

Metrics of Societal Benefits from Health Care Impacts

Institutional Realm of Society

Educational and scientific performances

- **Improved health care and research skill capabilities**
- **Recognition of Canadian medical science internationally**

Government and national performance objectives

- **Reduction in energy consumption**
- **Improved relations with the aboriginal population**

Economic performances

- **Reduction in sick days in employment and gains in productivity**
- **Increased trade balances for health equipment and supplies and creation of new organizations to manufacture health supplies**

Health and welfare performances

- **Increased equality in health care and duration of life by class and gender**
- **Decreased pension and welfare payments as a consequence of various kinds of disabilities and aging processes**

Another important side-benefit in the improvement of health care is the reduction of pensions and welfare payments for individuals who have disabilities and/or require extended institutionalized nursing care. Sometimes governments make deliberate decisions about preferring to increase disability pensions as a way of decreasing

unemployment, as did the Dutch government during the 1970s and 1980s but the reverse process is also true. Good health care can reduce pensions because of fewer disabilities associated with the aging process. Finally, contributions to improved political performances, improved equality, and lower welfare expenditures payments might be considered as a way of quantifying the intrinsic value of medical research to society. How does one aggregate across quite dissimilar societal performances within the same morbidity? The methodological problem of aggregation is much greater here because the units are so dissimilar. The simplest solution is to use percent change so that the units are standardized and can be combined. Even this solution poses some major problems because some of these performances are quite difficult to quantify except perhaps in surveys, e.g. the aboriginal opinion of the Canadian government.

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

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

Priorities in the Selection of Metrics.

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

On the basis of these criteria, the first fifteen health care impacts listed in Figure Two should the focus because these are easier to measure than those metrics listed under summary measures and contributions to the knowledge background and because these provide quicker feedback to policy makers. Within these first 15 health care impacts, one could further eliminate the speed and quality measures and perhaps also the prevention measures. The result is only seven metrics, ones that are fine-grained and can major incremental improvements obtained in research studies. Typically, one can ask for research projects to report exactly which of these health care outcomes has been studied. In each case, the corresponding economic gains can be relatively easily computed.

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

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

Aggregate benefits as ROI from medical research and computing social efficiency

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

Given the many stakeholders that are involved, I would recommend that the ROI on medical research investments be reported in the distinct categories outline in the Buxton and colleagues (1994) model: (1) health care impacts; (2) knowledge impacts; (3) increased capacity of the health care system; and (4) economic benefits. The inclusion of the society benefits might be made as well if deemed appropriate. Health care impacts have been divided into a number of metrics. Similarly, knowledge impacts have been divided into advances in treatment knowledge and contributions to the scientific literature. Increased capacity of the health care system involves both the increased capacity of health care personnel and of researchers. A number of economic benefits have been detailed because they are based on the health care impacts.

In the discussion of investments of Canadian dollars in distinct research areas, the idea was advanced to measure the social efficiency of these investments. My definition of social efficiency is an improvement in an output without the assignment of dollar value to it, which distinguishes it from economic efficiency or productivity. Thus, one

measures the social efficiency of the health care system at the macro level by using the changes in the age-gender population pyramid divided by medical expenditures as we did in our comparison of the health care systems of Britain, France, Sweden and United States (Hollingsworth, Hage, and Hanneman, 1989). This same logic can be applied to other kinds of health care impacts. For example, one might compute the amount of medical research dollars and researchers allocated to reduce the duration of treatment within a certain morbidity or allocated to increase the speed of post-treatment after an accident or stroke or other incident that requires rehabilitation. The advantage of having six research arenas is that it reduces the slippage between how medical research funds are spent and its consequences for the specific kinds of benefits detailed above, especially when the analysis is at the level of health care metric.

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

Conclusions:

Four Categories of Pay-Back

At the beginning of this white paper, a number of reasons were provided as to why it is important to focus on the treatment sector or meso level of analysis. The most important is the considerable variations between the technologies, procedures, and expertise needed to treat different morbidities. At this level, it has become possible to specify some twenty health care metrics (and of course more could be developed) that are

fined-grained enough to measure quite precise economic gains without having necessarily to measure QALYs, which usually are a consequence of a program of research that has unfolded in multiple countries over multiple decades. The meso level provides much more flexibility in deciding how many treatment sectors to be evaluated for determining ROI and which time periods to choose. The evaluator can assess only two or three sectors or twenty or thirty depending upon the priorities and largesse of the Canadian government. Similarly, the evaluator can choose a specific research study with a lag of several years or a very long stream of research over twenty or more years, except at we have noted, one should carefully assigned credit to the different countries and their researchers that were involved and evaluate the relative importance of their specific contribution, which is usually not done. Another degree of flexibility is that the evaluation might focus on only one or two metrics and study their potential health care impacts.

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

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

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

Knowledge production and capacity building

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

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

Capacity building has been treated in a number of ways. In the discussion of the actual benefits that accrue from the diffusion of treatment knowledge from new research findings, the health care personnel's human capital is up-graded. I have also suggested that this is an interesting way of measuring organizational learning. In the discussion of the investments in different arenas of research, metrics for measuring the improvement in the researchers skills have been suggested; admittedly these are indirect metrics because they do not actually measure how much is learned. In this context, the importance of participation in international research teams has been particularly stressed because it is

likely to increase the creativity of all the researchers and provide greater visibility of Canadian researchers and their achievements.

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

Informing policy

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

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

Another set of metrics deals with network gaps between arenas and with the lack of communication within complex research teams. While this kind of data collection is expensive, the "pay-back" of knowing how to increase the rate of scientific breakthroughs and also achieve better diffusion of knowledge advances may make the

cost worthwhile. In particular, this kind of evaluation can lead to better understanding of how to commercialization patentable products that flow from research and also to increase the rate of diffusion of knowledge throughout the health care system.

Just as I have argued that the detailed metrics of health care impacts allows for determining linkages with specific research studies, this list of the kinds of information feedback to policy makers will make it much easier to trace and measure when policy makers have been influenced and by how much from medical research. Furthermore, in the justifications for budget increases, one can also trace how these arguments shift from health care policy makers to the national party to parliament, indicating the values of the different stakeholders. These specific kinds of feedback thus become like a tracer bullet through the decision-making processes relative to medical research.

Health care impacts

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

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

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

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

Economic benefits

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

Societal benefits have been separated from economic ones because these are secondary outcomes that occur in time after the initial health care impacts and their economic gains. The logic of this exercise was to distinguish institutional realms with their own performances that can be affected by advances in health status. My intent was only to open the discussion about a broader range of impacts at the macro level. Some of

these might be of more interest to certain policy makers than economic gains, e.g. improvements in equality and in relationships with the aboriginal community.

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

1. Increases in knowledge about the health care problem
2. Advances in treatment knowledge;
3. Enhanced capacity in treatment knowledge among health care professionals;
4. Contributions to scientific knowledge;
5. Enhanced capacity in skilled researchers;
6. Economic gains from medical advances;
7. Societal benefits from medical advances.

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

Bibliography

Arnold, Eric 2004 "Evaluating research and innovation policy: world needs system evaluations" Research Evaluation, 13 (1), 3-17

Borbey, Peggy 2007 "measuring the impact of health research in Canada" conference on Economic Returns of Medical Research, Sigtuna, Sweden pdf at www.vr.se/huvudmeny/pressochnyheter/kalendarium/kalendariumarkiv/economicreturnsofmedicalresearch.5.5ec4ab911161ebf14dc80005402.html

Buxton, Martin 2007 "Estimating economic benefits at the national level" conference on Economic Returns of Medical Research, Sigtuna, Sweden pdf at www.vr.se/huvudmeny/pressochnyheter/kalendarium/kalendariumarkiv/economicreturnsofmedicalresearch.5.5ec4ab911161ebf14dc80005402.html

Canadian Academy of Health Sciences 2005 Developing a CIHR Framework to Measure the Impact of Health Research (unpublished document of the Canadian Academy of Health Sciences)

Canadian Academy of Health Sciences 2007 Prospectus for a Major Assessment: The Return on Investments in Health Research: Defining the Best Metrics (unpublished document of the Canadian Academy of Health Sciences)

Canadian Academy of Health Sciences n.d. "A Framework to Measure the Impact of Health Research" memoradum

Curran, Brenda 2007 "Impact assessment study from Irish health research" conference on Economic Returns of Medical Research, Sigtuna, Sweden pdf at www.vr.se/huvudmeny/pressochnyheter/kalendarium/kalendariumarkiv/economicreturnsofmedicalresearch.5.5ec4ab911161ebf14dc80005402.html

Hage, Jerald 1972 Techniques and Problems of Theory Construction in Sociology (New York: Wiley Inter-science)

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

Hage, Jerald and Jon Mote 2007 Transformational Organizations and Hyper Rates of Innovation: The Institut Pasteur, 1888-1918 (unpublished paper, Center for Innovation, University of Maryland)

Hanney, Stephen et al. 2004 "Proposed methods for reviewing the outcomes of health research: the impact of funding by the U.K.'s Arthritis Research Campaign" Health Research Policy and Systems, 2:4

Hollingsworth, J. Rogers, Jerald Hage and Robert Hanneman j1989 State Intervention in Health Care (Ithica, N.Y.: Cornell University Press)

Jordan, Gretchen 2006 "Factors influencing advances in basic and applied research: variation due to diversity in research profiles" in Jerald Hage and Marius Meeus (eds) Innovation, Science and Institutional Change: Handbook of Research (Oxford: Oxford University Press) pp. 173-195

Kolata, Gina 2008 "Drug said to cut risk of prostate cancer 30%" International Herald Tribune, 16 June, Monday: 4.

Lutkenhaus, Joe 2008 "Tinkering with Acellular Division" Science, 320, 9 May: 755-56

Marshall, Eliot 2008 "Pros and Cons of Mass Screening of High Risk Lung Cancer Patients" Science, 320, 2 May: 560-603

Miller, Greg 2008 "Tackling Alcoholism with Drugs Requires different Therapies for different Individuals" Science, 320 11 April: 168-170

Nelson, Richard (editor) 1993 National Innovation Systems: A comparative analysis. New York: Oxford University Press

Pelz, Donald and Frank Andrews 1976 Scientists in Organizations (New York: Wiley)

Robach, Kerstin and Per Carlson 2007 "Economic Value and Innovation" conference on Economic Returns of Medical Research, Sigtuna, Sweden pdf at www.vr.se/huvudmeny/pressochnyheter/kalendarium/kalendariumarkiv/economicreturnsofmedicalresearch.5.5ec4ab911161ebf14dc80005402.html

Stokes, Donald 1997 Pasteur's Quadrant (Washington, D.C.: The Brookings Institution)

Weisel, John 2008 "Engimas of Blood Clot Elasticity" Science 320 25 April: 456-57

Wooding, Steven 2007 "The payback model—an overview" conference on Economic Returns of Medical Research, Sigtuna, Sweden pdf at www.vr.se/huvudmeny/pressochnyheter/kalendarium/kalendariumarkiv/economicreturnsofmedicalresearch.5.5ec4ab911161ebf14dc80005402.html