CHSRF
CANADIAN HEALTH SERVICES RESEARCH FOUNDATION

FCRSS
FONDATION CANADIENNE DE LA RECHERCHE SUR LES SERVICES DE SANTÉ

HEALTH HUMAN RESOURCES PRODUCTIVITY: WHAT IT IS, HOW IT'S MEASURED, WHY (HOW YOU MEASURE) IT MATTERS, AND WHO'S THINKING ABOUT IT

FEBRUARY 2010

PREPARED FOR THE CANADIAN HEALTH SERVICES RESEARCH FOUNDATION AND THE MICHAEL SMITH FOUNDATION FOR HEALTH RESEARCH

ROBERT G. EVANS
DAVID SCHNEIDER
MORRIS BARER

www.chsrf.ca

Michael Smith Foundation for **Health Research**

Acknowledgements

We are grateful to Devon Greyson for undertaking the literature search, Colleen Cunningham for her assistance with the e-mail survey, Steve Morgan for collaboration in the early, design and information assembly stages of this project, and to Charlyn Black for assistance in interpretation of some of the evidence on the effectiveness of care. We readily absolve them of any responsibility for what appears on these pages. The leads provided by those who responded to our 'snowball' survey proved invaluable. We are grateful to all of those who gave so willingly of their time, expecting nothing in return (and getting precisely that), to help make this a more complete report.

This document is available on the Canadian Health Services Research Foundation web site **(www.chsrf.ca).**

1565 Carling Avenue, Suite 700
Ottawa, Ontario
K1Z 8R1

E-mail: info@chsrf.ca
Telephone: 613-728-2238
Fax: 613-728-3527

This research report is a publication of the Canadian Health Services Research Foundation. Funded through an agreement with the Government of Canada, CHSRF is an independent, not-for-profit corporation with a mandate to promote the use of evidence to strengthen the delivery of services that improve the health of Canadians. The views expressed herein are those of the authors and do not necessarily represent the views of CHSRF or the Government of Canada.

TABLE OF CONTENTS

KEY MESSAGES

- Improvements in productivity are the fundamental source of most increases in the material well-being of human populations. Although a relatively simple concept, productivity can be difficult to define, and changes difficult to measure in practice.

- In principle, health human resources productivity (HHRP) should be defined in terms of the relationship between health outcomes achieved (health status protection or improvement for individuals or populations) and the health human resource inputs (time, effort, skills and knowledge) required.

- The vast majority of current HHRP literature does not consider health outcomes, often using inappropriate and misleading measures of output. For example, more MRIs or more radiologists may contribute to increased procedural output but do not necessarily result in better health outcomes or improved productivity.

- Opportunities for increasing HHRP may be realized:

  - by examining unexplained variations in clinical practice evident in current comparative studies; and

  - through new ways of deploying health human resources that take advantage of full scopes of practice and roles, particularly within collaborative practice models.

- Many examples of HHRP-related successful innovation may not make their way into published literature because "getting published" is not a priority of those involved with HHRP-related innovation at the "coalface."

- Key points from this scoping review included:

  - Virtually all health services research is related in some way to HHR productivity, and yet few studies are explicit about their relevance to productivity. It would not be productive, or even feasible, to conduct a single summative synthesis of the literature in this area. Future research questions, primary and secondary, about HHR productivity need to be specific and sharply focused.

  - Most, if not all, HHRP literature focuses on inputs and outputs measured in terms of activities or processes rather than health benefits. The exception may be in the clinical variations literature, where there has been a clear focus on outcomes but little or no reflection on the implications for HHR productivity.

- Past research provides few clues as to when and why decision makers make improving HHR productivity (measured "properly") a priority. The impediments to productivity improvement include perverse incentives and misaligned objectives (e.g. other agendas take precedence). As a result, despite the fact that there are points of light scattered throughout the system, a large implementation gap persists between potential and actual improvements.

- Recommendations for further work in this area:

  - The identification and dissemination of success stories – where particular HHRP enhancements or interventions have led to improved health outcomes with similar or reduced inputs, or stable health outcomes with reduced inputs. The aim would be to find common success factors underlying such productivity gains.

  - HHRP prospective primary research; for example, evaluations of organizational, funding or programmatic changes or pilot initiatives that focus on the relationship between inputs and health outcomes/benefits.

EXECUTIVE SUMMARY

BACKGROUND

The healthcare sector makes up roughly one-tenth of the economic activity of modern economies, and labour inputs make up a large share of its costs, relative to other industries. As a result, the measurement, tracking and improvement of labour productivity in this industry, referred to here as health human resources productivity (HHRP), should be of significant policy concern.

In principle, HHRP should be defined in terms of the relationship between health outcomes achieved (health status protection or improvement for individuals or populations) and health human resource inputs (time, effort, skills and knowledge) required. However, the vast majority of HHRP literature defines HHRP as the ratio of procedural and service outputs over inputs measured in terms of numbers of personnel, or time.

The objectives of this scoping exercise were to prepare a "state of the science" report that includes:

- an overview of existing definitions and concepts of HHRP from healthcare and from pertinent non-healthcare domains;
- a summary of the important contributions on this topic in the scientific and grey literature, with an indication of the relative strength of the evidence;
- an overview of the leading researchers/centres with expertise on HHR productivity in Canada and elsewhere and current initiatives in policy and research (where available); and
- gaps and priorities for further research (syntheses and development of new knowledge) identified within the literature regarding practical concepts and definitions of HHR productivity for the current Canadian HHR planning and evaluation context.

METHODS

This scoping exercise on HHRP involved a structured, though non-exhaustive, search of electronic databases, along with a survey questionnaire to core (mostly Canadian) researchers and research institutions where work on HHRP was known or seemed most likely to be found. Virtually all research in the health services arena has some bearing on HHRP. This has meant that there was a vast potentially relevant literature; most of it turned out to be "low-grade ore" with a very low yield of relevant work. A snowball questionnaire strategy using a targeted questionnaire distributed by e-mail was far more productive in terms of surfacing research of relevance to HHRP. In addition, the authors' own research experience pointed to other literature that both the electronic and questionnaire searches failed to identify.

RESULTS

Virtually all health services research is related in some way to HHR productivity, and yet few studies are explicit about the relevance to productivity. The bulk of the literature on HHRP addresses productivity defined in terms of procedural or service outputs, not health outcomes. As a result, this literature does not provide information on how HHR changes affect individual or population health status protection or improvement. In contrast, other health services research such as randomized controlled policy trials, program evaluations, and particularly the literature on small area variations in health services use, while not focused directly on HHRP, often does provide information on outcomes with implications for HHRP.

Area variations research examines variations in patterns of clinical practice across different geographies or organizations. Typically, these variations have not been shown to be associated with any measures of patient need or health outcome. Such findings have powerful, if indirect, implications for HHRP, though this literature has not been examined through an HHRP lens.

For example, very large regional variations across Canada in inpatient use suggest further scope for reductions in inappropriate admissions and corresponding improvements in HHRP. Hall and Tu (2003) reported large variations in per capita admissions for major cardiac diagnoses. Rates in urban areas were much higher in the east than in the west, and were markedly higher outside urban areas. The national average is nearly twice that in Vancouver. These variations have not been explained by differences in population needs or health outcomes.

Similarly, Alter et al. (2008) find large differences in rates of cardiac intervention across regions in Ontario. Rates are uncorrelated with measures of population needs or health outcomes, but they do correlate with cardiologist supply, which varies among regions by a factor of two to one. This parallels findings in the United States, where regional variations in care use are found to be correlated with the supply of providers, and concentrated among those services that are provided largely at the providers' discretion.

Some advances in health services management, such as the development of clinical guidelines, implementation of new information technology and capitalizing fully on scopes of practice, could potentially improve HHRP. However, these approaches have faced, and continue to face, the classic political dilemma of diffuse benefits and concentrated costs. Productivity improvement would provide societal benefits, but would come at a cost for those involved in providing the associated services, with a significant threat to jobs and incomes. Generally for healthcare systems, the dominant mutually satisfactory strategy for individuals and organizations paid to provide care is to do more with more – more inputs, more outputs, and a shared belief that improved outcomes will result.

IMPLICATIONS

It would not be productive, or even feasible, to conduct a single summative synthesis of the literature in this area. Future research questions, primary and secondary, about HHR productivity need to be specific and sharply focused. There are examples of productivity improvement – some major – but they have not always been recognized as such and, even when identified, have been difficult to generalize. The field could be advanced through the documentation of "stories" that reflect successful HHRP enhancements. Such documentation, along with analysis of critical factors important to the success of those interventions, is a necessary foundation for more systematic review and, in time, more widespread uptake.

Further primary research on HHRP is also warranted, with a focus on the impact on patient outcomes of different approaches to health human resources deployment. This could include the evaluation of pilot initiatives or of programmatic organizational, funding or other changes, where there are opportunities to examine inputs (including HHR), alongside resulting outputs and outcomes. These might use time series, policy trial or other methods.

PREAMBLE – THE MANY MEANINGS OF PRODUCTIVITY

The concept of "productivity" is very simple in principle, but rather slippery to pin down in practice. At the most abstract level it is a relationship between one or more inputs to a production process (in economese, "factors of production") and one or more outputs from that process. The activity of most production converts **real resources** – human time and energy, raw materials, services of various forms of capital – into **commodities** that are of value to individuals or groups.[2]

The services of capital may flow from physical capital (buildings and equipment of various forms) or human capital (skills and knowledge learned and possessed by individuals or groups) or the intellectual capital of a society (know-how available, in principle, to everyone). This categorization distinguishes the time and effort that humans devote to a production process from the accumulation of skills and knowledge that guide their efforts. These latter represent a form of "capital" acquired through prior investment of time and effort. Productivity is then a measure of the volume of outputs achievable for a given volume of inputs. Improvements in productivity are the fundamental basis of all increases in the material well-being of human populations, from the Stone Age onwards, as they make possible the increased availability of valued commodities; crudely put, more bang for the buck.

The term "total factor productivity" is used to refer to the output achieved from some combination of all the different factors of production. If output is rising faster than the (weighted average of) inputs of both capital and labour, then total factor productivity is rising. Strictly speaking, however, if all the factors of production are identified and appropriately measured – the time, skill, knowledge and intensity of effort of labour input, the quality of raw materials, and all the dimensions of capital – it is unclear how total factor productivity could change. "Working smarter," or working with more advanced forms of equipment, for example, both imply increased capital in the form of intangible know-how. For our purposes, however, the more relevant measure is the partial factor productivity associated with labour, the amount of output generated per worker, or per hour worked at a given level of effort. Increases in productivity per worker, whether from more or better complementary physical capital, or greater skills and knowledge per worker, or from advances in the common possession of social know-how, are the basis for increases in commodities per person in a community – for a rising (average) standard of living.

The modern healthcare sector makes up roughly one-tenth of the economic activity of modern economies, and labour inputs make up a relatively large share of its costs, relative to other industries. It is thus understandable that the measurement, tracking and improvement of labour productivity in this industry, or if one prefers, Health Human Resources Productivity (HHRP), should be of policy concern.

The reference to "human resources" rather than the more traditional "labour" may be interpreted as reflecting the fact that what is of interest is the combination of both "raw" labour (time and effort) and embodied human capital (skills and knowledge).

SCOPE OF THE SCOPING EXERCISE

The scoping exercise on Health Human Resources Productivity (HHRP), the results of which are reported here, was intended to consist of three linked sets of activities: the gathering of information and research findings related to the topic, a review and analysis of the information gathered, and the production of a report in standard CHSRF 1-3-25 format. The first activity had, in turn, two parts: a structured, though non-exhaustive, literature search; and the identification and administration of a survey questionnaire to core (mostly Canadian) researchers and research institutions where work on HHRP seemed most likely to be found.

---

[2] It is extremely important to understand that money is NOT a resource. The monetary authorities in a modern economy can "print" money as desired; in and of themselves currency and bank accounts produce nothing. Rather, money may enable one to acquire the real resources that are productive. But if the resources are not available, no amount of money helps. A dollar bill might perhaps serve as a make-shift bandage, or a bundle of them help to start a fire for warmth, but that is probably the limits of the health benefits of money per se.

Going into this work, we anticipated that the literature might reasonably be sifted and sorted into a three-dimensional virtual storage cube. In one dimension of the cube, materials would be categorized by type of HHR: physicians; nurses; any/all other categories of HHR considered in isolation; or integrated teams and other forms of collaboration across personnel types. In a second dimension, items would be categorized as one of: HHRP concept/methods development; HHRP applied measurement and evaluation; HHRP reporting; or HHRP policy development/application. In the third dimension, works would be classified as involving either *process* measures of productivity or *outcome-related* measures of productivity. The task as we saw it at the outset was then to provide a general overview of HHRP definitions and concepts; a reporting of important contributions to existing HHRP research; an overview of research centres and expertise where HHRP research exists; and a summary of research gaps identified in the HHRP literature and through contacts with core research institutions.

The modern healthcare sector draws on a very wide range of human resources, from some of the highest levels of human capital in society to some of the most basic. The central role of human resources throughout this sector presents a problem for a scoping exercise. It is not possible to carve out some sub-set of health services research that is exclusively or predominantly HHRP; there, is in fact, no part of the whole field of health services research (HSR) that is beyond the scope of HHRP. Everything encompassed by HSR has implications, directly or indirectly, for HHRP[3]. A synthesis of the whole field of HSR from an HHRP perspective would be too large to be useful.

Sensitive, therefore, to the risks of dental disfigurement that result from biting off more than one can chew, or conversely the unhelpfulness of a one-line report – "It's everything" – we focus in this report on a number of sub-fields of HSR in which research is actively underway that we (and others) feel has a particular bearing on HHRP, or that has it as its primary rationale.

METHODS

In attempting to map the scope of research work in the field of HHRP, we employed three more-or-less complementary approaches, two of which are described in greater detail in the appendices to this document.

We began by specifying, in consultation with a professional librarian, a set of terms or keywords that were used to search a number of computerized databases that we believed to be most likely to contain relevant material. The list of these databases, and the keywords used in our search, are reported in Appendix A.

We then developed a list of all the persons in Canada who, we believed:

- might be engaged in research in this area
- might know others so engaged
- might have been so engaged in the not-too-distant past
- administer institutions where research relevant to HHRP might be being carried out by others.

We mailed out a short questionnaire to the people or organizations on this list and included a request for them to tell us of any other people they might know of working in this area. We also asked for copies of or links to materials that they considered seminal works in this field. There was a gratifying response rate of 76% to the initial survey. Through this "snowball" strategy we identified 93 articles and 104 experts; we ended up contacting, or hearing from, a total of 76 people. The list of people or

[3] Without intending to be exhaustive, we determined that the sub- and related fields of health technology assessment, clinical practice guideline development, cost benefit/effectiveness/efficiency analysis, nurse staffing (on which there is a copious literature) and healthcare sector staffing models more generally, frontier efficiency measurement, stochastic frontier analysis, operations research, clinical practice variations, lean production, and inter-professional teamwork, all had something to say about, and were motivated at least in part, by an interest in improving productivity in the healthcare sector through more efficient use of health human resources.

organizations to whom/which the original questionnaire was e-mailed, the questionnaire, the covering letter, and the letter of thanks to respondents are all contained in Appendix B to this report.

Finally, we added a considerable amount of material from our own past experience in the area in the process of trying to organize and bound the field of HHRP. We have gone far beyond simply cataloguing the related works that we were able to identify, with the result that this report is a combination of traditional scoping output, theoretical and framework development, and analysis and interpretation of the "field" as we find it today. We did try initially to stay within the concept of scoping, but we did not succeed.

Appendix C provides a comprehensive list of the various materials associated with this report, in four sections. C.1 contains the references for the main text of the report, while C.2 identifies what we regard as the key papers from those references. C.3 and C.4 list the references culled from the computerized search and from the snowball process respectively.

As discussed in more detail in Appendix B, the snowball strategy was far more useful than the keyword-based literature dredge. The computer search turned up far too many references for us to handle (5,258), and when we tried to flag them as relevant or not, and to map them into an organizing framework, we found that our inter-rater reliability was weak. There is simply not enough information in titles alone, and even in the minority of cases for which abstracts were available, the abstracts were also often insufficiently informative to enable us to judge their relevance. Going beyond that, to assemble and then read 5,000 articles was well beyond the intent of, and resources available for, this or any other scoping exercise.

Even more profoundly, we found from the snowball that our initial keyword selection had left out some critical terms. As we noted above and document below, virtually all of the research in the health services arena has some bearing on HHRP, and there is, in addition, research outside HSR, and even outside the health arena, that is relevant. Any attempt to bind the task using keywords is doomed to result in missed works of relevance. In addition, several different terms may be used for the same concept, depending on the discipline and field of study.

We tried to pick up missing items through a broader literature search using an expanded keyword list; however, in applying this expansion to a single database, we discovered that we might well be faced with as many as 12,000 references. At this point we admitted defeat and turned our focus almost exclusively to the results of the snowball survey. In this case at least, the snowball strategy was far more productive than any keyword-based literature search strategy that we could design.

REFLECTIONS ON MEASURING THE PRODUCTIVITY OF HEALTHCARE

In principle, the study of productivity should be as simple as the concept itself – add up the outputs and the inputs, divide the former by the latter, and observe the behaviour of this ratio over time, across regions, or between different production sites or alternative techniques of production. For the study of HHRP, the denominator can be restricted to (all or particular) HR inputs. In practice, however, there are significant technical and conceptual problems in the measurement of productivity in general, and these are compounded by the special characteristics of healthcare.

The technical problems arise because production processes almost always require multiple types of factor inputs, and frequently generate several different forms of output simultaneously. Measuring productivity then requires the creation of aggregate indexes for both inputs and outputs. The appropriate weights to be used for constructing such indexes have been the subject of considerable theoretical study; the question is far from trivial. But further, these theoretical analyses have typically proceeded on the pivotal assumption of competitive private markets in which prices of inputs reflect their opportunity

costs and prices of outputs reflect their value (defined in terms of willingness to pay) to informed consumers. These conditions do not, and more fundamentally cannot, in general hold in the healthcare sector, anywhere in the world.

The more serious problem, however, is that what is valued from the activities of the healthcare sector is not the traditional concept of outputs – commodities – but *outcomes,* improvements in or at least preservation of health status for some individual or group. With minor exceptions, the goods and services produced in the healthcare sector – diagnostic and surgical procedures and drugs of all types, visits and consultations – are "bads" that in themselves make the recipient uncomfortable in a variety of ways; most of us would prefer to use less of them, not more. They are undergone, consumed and endured in the expectation that they will have consequences in the form of health status improvements or protection that are of significant value, or at least of sufficient value to compensate for the inconvenience, discomfort, distress, and sometimes risk and fear associated with them.

That expectation is frequently justified; no one should discount the achievements of modern medicine. However, that merely reinforces the fact that the outputs of the healthcare sector, traditionally measured, are not in themselves *desiderata.* When the productivity of that sector is evaluated, as it all too frequently is, in terms of quantities of particular outputs of goods and services – patient visits and consultations, hospital bed-days, procedures of all kinds, the panoply of drugs consumed – a very powerful assumption is being made that these outputs always and necessarily generate corresponding benefits in terms of health status. This comfortable presumption is simply not supported by the available evidence (e.g. Fisher et al. 2003a,b; Alter, Stukel and Newman, 2008; Goodman and Fisher, 2008; Goodman and Grumback, 2008). To the extent that the healthcare sector absorbs resources and generates outputs that do not result in improvements in health outcomes or other measures of well-being, then any measure of productivity that includes these forms of output will be over-stated and highly misleading. An institution or system that is becoming increasingly efficient – more output per unit of input – at producing ineffective or even harmful care is not becoming more productive in any meaningful sense.

The American financial sector offers a clear illustration of the point, as well as showing that the issue is not confined to healthcare. Senior executives in the financial products division of the (formerly) giant insurer AIG apparently had a contractual, legal right to a total of $165 million in bonuses for the year 2008. We have not been privy to the "trigger" details, but a plausible explanation is that these bonuses were calculated on the basis of the revenue that the executives brought into the company. By this measure they were collectively highly productive. In the process, however, they grossly miscalculated the risks involved, and drove the company into bankruptcy. It remains in business sustained only by huge financial subsidies from American taxpayers. A similar story of massive bonuses tied to defined performance measures and preceding a massive bankruptcy is observed in the collapse of Enron (Ackman 2002; Andrews & Baker 2009).

Outputs versus outcomes raises thorny questions in healthcare, with obvious political dimensions because in this sector people are being paid for producing services whether or not there is any systematic investigation of their effect on health outcomes. Any divergence between outputs and outcomes is typically far less obvious and dramatic than in the cases of AIG or Enron. In the public discourse the assumption that whatever is provided must have been "needed" is rarely challenged.

These political issues may explain why, despite the fact that modern public health insurance systems have now been in place for more than a generation, so little serious effort, or at least so little progress, has been made to measure their productivity. A naïve assumption that there is a widely shared interest in improving productivity in the healthcare sector, and that all that is lacking are better data and tools

of analysis, would be seriously misleading.[4] This will, we believe, emerge at several points below. (See also Evans, 2008 [*i.e.*1999], attached as Appendix D.)

Any measure of productivity in the healthcare sector must nevertheless start from the fundamental understanding that it is health outcomes, improvement, or at least maintenance of the health status of individuals and communities that constitute the purpose and the justification for all health sector activity. Inappropriate care may yield no benefit whatsoever, and in some cases can do serious harm. The peculiar institutional and regulatory structure of the sector in all modern societies reflects that fundamental reality, and is based on the need to protect patients and others from the consequences of inappropriate care. That it does not do a perfect job should by no means be taken as suggesting that it is without powerful beneficial effect!

INPUTS, OUTPUTS, OUTCOMES, AND THE NATIONAL ACCOUNTS

The bulk of the literature on HHRP, alas, addresses productivity defined in terms of outputs, not of outcomes. In the light of the evidence from those studies that do address outcomes, this represents a serious limitation, as we will see below. But first, there is a small but potentially influential research program that begins even farther "from the coalface," not only inferring outcomes from outputs, but even inferring outputs themselves from inputs. Sharpe and Bradley (2008) assert:

> "The current approach to the measurement of health care sector output in Canada relies largely on the use of volume of inputs to the health care sector as a proxy for volume of outputs."

On these assumptions, there can be no change in overall measured productivity. If productivity refers to a relationship between inputs and outputs, and if the latter are defined by the former, then their relationship is fixed by assumption; hence Sharpe and Bradley's emphasis on the need for more appropriate measures of health sector output independent of input measures, and for appropriate measures of relative prices to be used in compiling aggregate indexes of the volume of sector output.

Their concern is with the consistent integration of the healthcare sector into the National Accounts framework, and the extent to which these accounts may underestimate the growth of productivity as conventionally measured. It would certainly be possible to develop measures of healthcare sector output and "productivity" on a basis consistent with the National Accounts framework. Indeed, the final report of the Atkinson Review (2005), which addresses the much broader question of the treatment of government services in the National Accounts of the U.K., includes health services. That review, and associated commissioned research (Dawson et al., 2005), provides a comprehensive discussion of the technical problems involved, and makes recommendations as to how to deal with them, including the necessary expansion of the statistical sources.

It is important to understand, however, that the National Accounts provide a measure of economic activity, not of economic or any other dimension of welfare. Measures such as the Gross Domestic Product (GDP) are routinely used, particularly in the media and in political discourse, as if they were an aggregate measure of national well-being. More commodities, of whatever type, are assumed to equal more happiness; goods are, by definition, good. However, this identification is erroneous; the GDP is at best an approximation to any measure of well-being, even outside the peculiar context of

[4] These political difficulties have a long history. A classic example is provided by the Boston surgeon E.A. Codman, whose determined efforts in the early 20th century to establish the systematic study and review of the "end results" of all hospital care foundered in the face of the intense hostility of his colleagues (Donabedian, 1989).

healthcare. Students of national accounting have always been very explicit and clear that these accounts measure only economic activity and by extension capacity. Measuring well-being is quite another matter, and much more complex.

The Atkinson Review is quite explicit about the ultimate relevance of health outcomes rather than simply healthcare outputs:

> "**8.24** The current methods capture *activities* carried out. Under Principle B (see paragraph 4.24), an output measure should be adjusted for the attributable incremental contribution of the activity to individual or collective welfare.This should include capturing any change in outcomes which is attributable to the use of the inputs. A basic count of activities does not measure the quality of the output such as change in quality of patient experience or clinical effectiveness. This is a continued weakness of the current method …" (Atkinson, 2005)

The choice of the word "quality" as a general label for what is here identified as "patient experience or clinical effectiveness" is, however, quite unfortunate given the protean nature of the concept of "quality." Following Donabedian's (1966) now-classic distinction between *structure, process,* and *outcome* measures of quality, researchers and particularly regulators have over the years tended to give lip service to the importance of outcomes, and then to focus on structure and/or process measures of quality. This can provide an automatic justification for lower productivity as higher quality.[5]

Improving the measures of healthcare system activity or outputs is surely an objective to be applauded, but from the perspective of healthcare policy, the results are not without risk. The extensive discussion of outcome measurement in Atkinson (paragraphs 8.46-8.66) does not bring us any closer to integrating outcomes into the National Accounts. Indeed, as emphasized above, this integration would in fact be inconsistent with the conventions of those accounts.

National accountants have not been insensitive to the importance of distinguishing health outcomes from healthcare, as illustrated by the discussion in Atkinson (2005) and in United Nations Statistics Division (1993):

> "16.136. The output of health services needs to be clearly distinguished from the health of the community. Indeed, one reason for trying to measure the output of health services may be to see the effect of an increase in the volume of health services on the health of the community. This obviously requires a measure of the volume of health services that is different from health itself. It is well-known that there are many other factors such as sanitation, housing, nutrition, education, consumption of tobacco, alcohol and drugs, pollution, etc., whose collective impact on the health of the community may be far greater than that of the provision of health services."

However, giving greater concreteness and precision to measures of the outputs of the healthcare sector and enshrining them in the National Accounts, where the principle that "more is better" rules, could yield results that are not merely irrelevant but misleading, potentially actively mischievous, and in the extreme, dangerous to the health of individuals and populations. Increased production of ineffective or harmful healthcare goods or services will be counted as "more output" by the conventions of the National Accounts, trumpeted as contributing to growth in the GDP, and correspondingly included in the numerator of "productivity" estimates.

---

[5] This tendency reached its nadir in the work of an otherwise highly distinguished economist who once *defined* quality by the volume and cost of inputs used per unit of output, in effect as the inverse of productivity (Feldstein, 1974).

In high-income countries, international differences in aggregate population-level health status bear no apparent relationship to differences in levels of healthcare provision. The aggregate correlations are very difficult to interpret; the most that can be said is that among those countries there is no apparent relationship between health spending, or the supply of particular types of personnel, and any measures of health outcome (e.g. most recently: Wagstaff, 2009; Watson and McGrail, 2009). Apart from the technical problems of disentangling prices and quantities in a consistent way across different national healthcare systems, the two major impediments to finding any such relationship, probably insurmountable, are the aggregation problem and the attribution problem.

The aggregation problem arises from the difficulty of finding some aggregate index of health resource inputs or of health system activity, and a corresponding aggregate index of health status. The attribution problem arises from the fact that as noted above (United Nations Statistics Division, 1993) healthcare, of whatever form, is only one of the many features of a society that contribute to population health, and most of these other factors are not well measured, if at all. One should not therefore expect to find any systematic relationship between aggregate measures of health status and measures of healthcare input; however, they may be aggregated. Indeed, in some cases the measures of health status themselves are radically inconsistent. The United States, for example, scores relatively badly on most measures of population health, such as life expectancy or Potential Years of Life Lost (PYLL), yet Americans' self-reported health status is among the highest in the world. Japan, with the world's best mortality results, self-reports very low health status (Organization for Economic Co-operation and Development, 2009).

Attempts to deal with the attribution problem have led to the definition of a sub-set of "medically amenable conditions" – causes of morbidity or mortality for which effective therapies exist (OECD *op.cit.*; Nolte and McKee, 2008). The rates of mortality from such medically amenable conditions might then be expected to be related (negatively) to the availability of medical personnel. One such study has found that Canada, with one of the lowest rates of physicians per capita among high-income countries, nonetheless has among the lowest mortality rates from a recently-defined set of medically-amenable conditions (Watson and McGrail, *op.cit.*). Both the definition of medically amenable conditions and the aggregate results are, however, problematic and potentially contentious.

Several recent papers have compared one or more high-income countries with the United States, on the grounds that levels of health expenditure are so much higher there, both in absolute terms and relative to GDP, than in any other high-income country. If Americans receive so much more healthcare, might one not expect to see differences – in particular, more narrowly defined – areas of healthcare access and health outcome? As noted above, the comparisons are made difficult by the fact that the relative prices of healthcare goods and services are significantly higher in the United States, compared with the prices of other commodities, and the mix of measures of capacity, utilization, and outcome is highly variable across countries.

The consistent message that emerges, however, is that in advanced countries there is no evidence that more activity, in aggregate, correlates with better health outcomes. Indeed, in the case of the United States, considerably more activity relative to the United Kingdom is actually found to be associated with considerably worse health outcomes (Banks et al., 2006).

IMPLICATIONS OF THE "CLINICAL VARIATIONS" EVIDENCE FOR HEALTH HUMAN RESOURCE PRODUCTIVITY

The limited data available on health outcomes, however, and the difficulty of linking aggregate outcomes to healthcare activity, are reflected in the fact that, as noted above, by far the preponderance of explicit HR productivity research in healthcare focuses on the relationships between inputs and outputs, not outcomes. More simply yet, health human resources productivity (HHRP) is commonly measured by some single or compound measure of output, *per* some measure of input of personnel time. Thus, one

may count average patient visits per full-time equivalent (FTE) physician, or hours of nursing care per hospital patient day, or prescriptions filled per pharmacist per working day. The appropriateness of the care is inferred implicitly from the observation that it was provided – in effect not considered at all.

An exception to this generalization, with powerful and quite disturbing implications for our understanding of HHRP, is the study of clinical variations. "Clinical variations" is short-hand for the large variations in patterns of clinical practice among countries, small or large geographic regions, or academic medical centres, that cannot be shown to be associated with any measures of patient need or health outcome. These have been extensively documented over decades and in many countries.

The leading centre for such research is the group at Dartmouth University, under the leadership of John Wennberg and Elliot Fisher, but there is a long tradition of such observations in Canada, going back to work by Eugene Vayda in the 1970s (Vayda, 1973; Vayda et al. 1976; Stockwell and Vayda, 1979), and subsequently by the team led by Noralou and Les Roos at the Manitoba Centre for Health Policy. More recently David Alter, Therese Stukel, Jack Tu and others at the Institute for Clinical Evaluative Sciences (ICES) in Toronto have studied this phenomenon in Canada (see, for example, Tu et al., 2006; Alter, Stukel and Newman, 2008). All these researchers have uncovered remarkably large interregional variations within and across provinces or internationally, in use of services – variations unrelated to evidence of corresponding variations in either need or outcomes. Interpretation of these observations has long been hampered by the difficulty of ruling out the possibility that the variations are linked to unobserved differences in patient needs or outcomes achieved – in research design terms the problem of "missing (explanatory or effect) variables" – in more basic terms "My patients are different." The onus, the burden of proof, has until recently been placed implicitly but squarely on those suggesting that observed variations are not clinically justified. But in the last decade major advances have for the most part addressed these limitations. The variations remain intact, and the excuses have worn thin.

Breakthrough articles by Fisher et al. (2003a, b), published back-to-back in the *Annals of Internal Medicine,* using very large databases and advanced techniques for standardizing for differing patient needs, exposed very large differences across the 306 defined hospital service areas in the United States in the frequency of provision, and corresponding costs, of various services and procedures among the Medicare (over-65) population. These differences in the average intensity of servicing received by different populations were associated with small but highly significant differences in patient outcomes, at least as measured by mortality. However, more intensive servicing was associated with *higher,* not lower, mortality, nor were these higher risks compensated by any differences in (self-reported) patient satisfaction.

More intensive rates of servicing were also closely associated with greater availability of hospital facilities and personnel. This finding is particularly disturbing for the analysis of HHRP, because it implies that where more personnel are available (per capita, age- and need-adjusted) more services are provided, but health outcomes are, if anything, worse. Productivity measures focused on the relationship between outputs of visits and procedures might show productivity rising or falling (or flat) depending upon whether activity volumes rose more, less or in proportion to the increased availability of personnel and other forms of capacity. Regardless of the relationship between resources and outputs, however, on the measures that matter – health outcomes – the impact of more personnel is in aggregate unambiguously negative.

Subsequent analysis by type of service has shown that the more intensely serviced areas in the United States do not receive more services of types that might be considered essential on medical grounds, nor of those discretionary services that might be considered reasonable for the patient to make the choice. Rather, the variations among geographic regions are most marked for those *supply-sensitive* services that are not dictated by hard evidence of effectiveness (or lack of it) but are at the provider's discretion (Wennberg et al., 2002): the greater the HHR capacity, the more physician-discretionary services per capita. "Standards of care" are then set by local convention, in response *inter alia* to the availability of resources.

Anyone familiar with the many-decades-old literature on supplier-induced demand might wonder why this would be a surprise. Still, the implications for HHR productivity are sobering.

Could quality differences explain the variations – more personnel and facilities, higher intensity of servicing, but higher quality of care? In fact, similar clinical variations have been found among very highly respected academic medical centres (Fisher et al., 2004; Wennberg 2005). And a widely used scoring system for quality of care has been found to correlate negatively, across states, with the availability of specialists per capita – though positively with the availability of general practitioners (Baicker and Chandra, 2004).

More recently, Goodman and Fisher (2008) explored the relationship among regions between physician supply and measures of both quality and patient satisfaction. The (age-sex adjusted) number of physicians per 100,000 varies widely across the 306 hospital service areas defined for the United States, from 271.8 in the highest quintile to 169.4 in the lowest – a difference of about 60%. Per capita supply in the median quintile was 204.8. However, the authors found no difference in patients' self-reported satisfaction with access to care, quality of care or their doctors' concern for their overall health, nor do Medicare composite quality scores show any difference between high- and low-supply areas. There is simply no indication that more doctors result in better care, even for very large differences in supply. American medicine appears to be on the "flat of the curve" (or below it).

It should not be inferred from these findings that physicians are uniformly hazardous to the health of populations. Goodman et al. (2002) studied the relationship between regional variations in neonatal death rates in the United States and the availability of neonatologists and neonatal intensive care beds, and found that there was a significant (negative) relationship between mortality and the number of neonatologists per 10,000 births (though not with bed availability). However, this was true only for the quintile of regions with the very lowest availability.

Between regions with an average of 2.7 neonatologists per 10,000 births, and the next lowest quintile with 4.3, there was a significant decline in mortality (odds ratio .93%). From there on up, from 4.3 to 5.9 in the medium quintile to 11.6 in the highest, there was no additional gain. The implication is that while there are regions that are under-provided, at least on this indicator, the overall supply is far above needs and severely mal-distributed. More would not help.

The term "flat-of-the-curve" medicine is often applied to this pattern, in which there are measurable health benefits when services per capita are increased from very low levels, but where further increases in services, or in this case neonatologist capacity, yield no further measurable benefit. If current levels of provision are on average far above the point at which further benefits cease to be observed, the system is operating on the flat of the curve.

Fisher (2007) sums up the American evidence:

> "*What I know:* Higher spending across regions and physician groups is largely due to overuse of *supply-sensitive services* – hospital and ICU stays, MD visits, specialist consults; *and – at the margin – more is worse.*
>
> *What I think I know:* Overuse is largely a consequence of reasonable differences in clinical judgment that emerge in response to local organizational attributes (capacity, clinical culture) and financial incentives that promote unnecessary growth and more care."

How far this generalizes beyond the United States is unclear, but similar findings for Ontario were reported by Alter, Stukel and Newman (2008).

> "Regional per capita cardiologist supply varied more than twofold across regions, but was inversely related to the regional cardiovascular disease burden." … "Residents in areas with more cardiologists were more likely to receive some form of cardiac intervention." … "However, the intensity of provision of cardiac health services was unrelated to regional cardiovascular disease burden and was not associated with improved survival." (p. 187)

This implies that increasing the numbers of cardiologists available to serve the Ontario population might be quite "productive" in terms of procedures performed, but the marginal productivity in terms of improved health outcomes is likely to be zero – or worse.

In fact, the whole field of diagnostic imaging in Canada displays similar variations and raises similar questions. The report *Medical Imaging in Canada 2007,* released in 2008 by the Canadian Institute for Health Information (CIHI), pulls together a large number of scattered and fragmentary pieces of information on time trends and regional variations in imaging capacity and use. It is difficult to assemble a comprehensive picture, but several main points do emerge.

First, imaging capacity in Canada is growing rapidly (CIHI, 2008, Figure 13, p. 38). In the four years from 2003 to 2007, the numbers of CT scanners and MRI machines grew by 29% and 49% respectively, from 325 and 149 to 419 and 222 (in 1990 the corresponding numbers were 198 and 19). These numbers may understate the growth in capacity insofar as the newer machines can produce more as well as more detailed scans, although in fact the total number of scans reported (by fiscal year) rose roughly in parallel with the numbers of machines – 32% and 47% (Table 5, p.91).

Despite this greatly expanded imaging activity, the number of imaging professionals per capita in Canada has remained roughly constant (Figure 63, p. 118). There has thus been an unambiguous and large increase in HHRP *as measured by outputs or activities per unit of input.*

Compared with other OECD countries, Canada was below the median in equipment per capita in 2005 (Figures 39 and 40, pp. 78 and 79), but data for a subset of countries indicate that Canadian imaging machines are more productive than those in several other countries, so the number of procedures per capita may be at or above the median (Table 5, p.109).

The comparison of Canadian numbers with international averages has received considerable attention in the rhetoric about the inadequacies of the Canadian healthcare system and the (successful) arguments for greatly increased funding. Less noticed, indeed studiously ignored, have been the data on international *variations* in capacity. Japan had 92.6 CT scanners and 40.1 MRI machines per million population in 2005; the Netherlands had 5.8 and 5.6 (Figures 39 and 40). The United States had 45.3 and 26.6; Germany had 15.4 and 7.1. Canada, at 12.1 and 6.1, was just below the medians of 14.7 and 6.9 (though helped somewhat by reporting 2006 data). However, there is nothing that one could call an "international standard," only mechanically computed averages. In these circumstances, to try to "keep up with the rest of the world" is to chase a Chimera.

What then do international data mean in the context of such radical dispersions? What they mean is that, once again, we are observing wildly variant patterns of clinical practice, wholly unrelated to any evidence of a connection between diagnostic activity and patient needs or outcomes. In fact, the data assembled for CIHI (2008) show similar large variations across Canadian provinces in both capacity and use. Alter, Stukel and Newman (2008) were able to combine regional-level data on diagnostic capacity and use with data on needs and outcomes; we do not have similar data for diagnostic imaging generally, but we have no reason to believe that, if we did, the story would be any different.

While the most extensive and carefully studied data on clinical variations emerges from the United States, and in particular from the research group at Dartmouth led by John Wennberg and Elliot Fisher, it would be quite misleading to imagine that this phenomenon is restricted to the United States. The international comparisons of diagnostic imaging show truly remarkable variations among high-income countries with modern medical systems.

The studies referred to above are only leading examples of findings of large regional variations within Canada; see also the discussion below of variations in inpatient admissions for cardiac care (Hall and Tu, 2003). There is much more of this analysis that could and should be done in Canada, and the documentation of these clinical variations offers an obvious way to focus efforts to improve HHRP.

More generally, while the United States is unique in its extraordinarily high levels of medical spending, the bulk of the difference between the United States and Canada is in inflated procedural and service prices and in administrative waste from private insurance. As discussed in more detail below, if American expenditures were adjusted for the extraordinarily high bureaucratic overhead costs generated by a private health insurance system, and for the much higher service and procedure prices in the United States (relative to Canada, or indeed anywhere else in the world), the share of American national income devoted to paying for healthcare itself would be very similar to that in Canada, and possibly even lower.

Evans (2007) offers a global comparison of costs and outcomes, while Woolhandler et al. (2003) estimate huge excess administrative costs generated both in the insurance system itself and in hospitals, medical practices, and other care settings by the requirements of a private health insurance system. Eisenberg et al. (2005) and Antoniou et al. (2004) detail, for coronary artery bypass grafting and total hip arthroplasty respectively, the total procedural costs on either side of the border. For both procedures the costs were nearly twice as high in the United States. For neither were there any differences in outcomes.

The average number of physicians per capita does not differ greatly between the two countries, and the Canadian physician-to-population ratio is close to the centre of the range reported by Goodman and Fisher (2008) for American regions. There is unquestionably a significant overlap between regional service patterns in Canada and in the United States (Lasser, 2006). It follows that similar results, if perhaps less extreme, would almost certainly be found here.

In the face of such observations, it is simply not credible to assume that more personnel are necessarily associated with better healthcare outcomes, whether or not they are associated with increased outputs of clinical services or other forms of activity. This highlights the importance of research programs at ICES and elsewhere that focus on the linkages between the availability of health facilities and personnel, the frequency of provision of services, the underlying burden of illness, and the consequences (if any) in terms of identifiable health outcomes.

The productivity of health insurance

The very large excess administrative costs of private health insurance in the United States, relative to the costs of universal plans in Canada and Europe, running in the hundreds of billions of dollars (Woolhandler et al., 2003), illustrate from a different perspective the contrast between outputs and outcomes. The outcome sought by purchasers of insurance is security – relief from risk. The policy is just a means to that end. However, most non-elderly Americans have in general far less security with respect to healthcare costs than citizens of other high-income countries. Quite apart from those entirely excluded from private health insurance, a high proportion of those with private coverage are in fact seriously underinsured against any major illness or injury (Schoen et al. 2005), and face significant restrictions on their employment mobility (Madrian, 1994). So they pay much more, and get much less, because so much of the costs of the private insurance industry are absorbed by activities to *avoid* paying claims (Geyman, 2008).

This massive waste in the private insurance industry itself represents only about half of the excess overhead costs generated in the process of propping up the private insurance industry. The other half are incurred in by individual physicians, hospitals, clinics, and long-term care facilities. These are the extra administrative activities necessitated by dealing with a fragmented insurance system ever alert for ways to limit its claims liabilities. And these costs do represent reduced productivity in the healthcare sector – the costs borne by healthcare professionals and institutions in order to be reimbursed for the services they have provided. In some cases the same consultant firms are selling software and advice to both sides – to insurers to help them avoid payment, and to providers to help them get paid – in effect, arms dealers in an administrative arms race (Evans, 1995; Geyman, *op. cit.*). All this activity is duly recorded in the National Accounts as economic output, and people are paid for it (often quite well), but it is a very inefficient way to provide either health insurance or healthcare. And, insofar as it restricts access to necessary care and severely increases uncertainty and economic strain, the gross inefficiency of the American insurance system has itself a major negative impact on health outcomes.

> "The best estimates suggest that something like 18,000 unnecessary deaths take place each year just because of inadequate health insurance. That's the equivalent of six 9/11s every year." (Krugman and Wells, 2006)

READING OTHER TEA LEAVES FOR PRODUCTIVITY MESSAGES – RCTS AND HTA

At the other end of the scale from the population-based clinical variations studies are those that focus on the efficacy of particular interventions. The huge literature reporting outcomes of randomized controlled trials (RCTs) is fundamental to HHRP in the direct sense that employing people to perform activities that can be shown to have no health benefit is by definition unproductive, while performing beneficial activities is. Insofar as the results of successful trials can be carried forward into clinical practice, and the neutral and negative results kept out of clinical practice, they can be thought of as outcome-focused productivity-enhancing activities.

The RCT literature, in turn, feeds into the studies under the general label of health technology assessment (HTA), or still more generally program evaluations of all types. If an intervention is without (positive) effect on health, then it deserves no further analysis. But an effective intervention still requires evaluation, of its range and context of benefit, and of the balance between benefits and costs in particular settings, relative to feasible alternatives (including doing nothing). Since all interventions in the healthcare sector involve some use of HHR, the outcomes of interventions represent part of the measure of the productivity of the personnel thus employed. Low-cost, high-benefit interventions are likely to be a highly productive use of the associated personnel. High-cost, low-benefit interventions, and no-benefit interventions at any cost, are correspondingly an unproductive use of HHR.

The RCT literature focuses exclusively on outcomes, seeking yes-or-no answers (within the limits of statistical analysis) to the question of effectiveness. The HTA process sometimes combines assessment of the evidence of effectiveness with a judgment as to how much effect is worth paying for. There is also a branch of literature that might be included as HTA that evaluates the impact of changes in the

nature or quality of complementary resources, particularly capital equipment, on HHRP. These studies may or may not deal with outcomes. The rapid and costly proliferation of high-technology diagnostic imaging in Canada, for example, has been associated with steady growth in the quantity and quality of imaging, but also in increased productivity of technical staff – in terms of outputs. There does not appear to be any evidence of improved outcomes.

Unfortunately, there is a strong bias in uptake of the findings of RCTs or HTA, with interventions that yield positive results being relatively rapidly taken up, particularly if they involve the deployment of new, more expensive, products, capital, or services. At the same time, interventions found to have no significant benefit or minimal benefits for relatively high cost tend, if already in widespread use, continue in use.

A leading example in the Canadian context is that of the justly celebrated "Ottawa Ankle Rules," a procedure that renders radiography unnecessary for nearly all cases of possible ankle fracture. These rules offer a significant improvement in HHR productivity, for a relatively common presenting complaint (Stiell et al., 1992; Stiell et al., 1994). Ten years after this ground-breaking research was reported, Bachmann et al. (2003) concluded, from a meta-analysis of 32 studies and 15, 581 patients, that:

> "Evidence supports the Ottawa ankle rules as an accurate instrument for excluding fractures of the ankle and midfoot. The instrument has a sensitivity of almost 100% and a modest specificity, and its use should reduce the number of unnecessary radiographs by 3040%." p. 1157

However, those potential gains have been squandered by clinicians. For example, while approving the rules in principle and claiming to follow them, emergency physicians have been found to be continuing to order procedures that application of those rules show to be of no value – "just in case." (Brehaut et al., 2005). Meanwhile, the developers of the Ottawa Ankle Rules have gone on to develop simple and reliable observational tests for several other forms of suspected fractures, potentially obviating the need for still other forms of imaging (Perry and Stiell, 2006). If physicians are equally incapable of understanding and appropriately applying these additional "low-tech" diagnostic tests, then the frequency of unnecessary imaging procedures and the extent of resource waste has presumably expanded.

RCT results can be linked with the clinical variations findings described earlier in an effective way, as shown by Skinner et al. (2006). For the United States, casual correlation of the growth over time in national spending on cardiac care with cardiac mortality appeared to show significant benefits, sufficient to justify the increasing costs (Cutler and McLellan, 2001). But analysis by Skinner et al. of the RCT findings on the relative effectiveness and costliness of alternative therapeutic approaches, combined with data on the differing mix of approaches in different regions, undermined this conclusion. The devil, as always, is lurking in the details.

The options available for the treatment of cardiac disease resolve into two more or less distinct clinical styles. These have been shown to be equally effective in achieving health outcomes, but are very different in their total costs, and the therapeutic choice is at the clinician's discretion. The observed large differences in costs among regions reflect the differing choices being made by clinicians in each region.

Higher-cost styles of care may or may not be technically efficient in their use of resources – inputs – to produce the procedures and services that they involve. However, their extra cost, arising from greater intensity of servicing and corresponding greater resource use, represents unequivocally lower productivity in the achievement of health outcomes.

RISING UTILIZATION, FALLING PRODUCTIVITY?

The link between high costs, over-servicing, and physician behaviour was dramatized for a mass audience by Atul Gawande's essay in the June 1, 2009 issue of *The New Yorker* (Gawande, 2009). The essay focuses on the Texas border town of McAllen, where the American Medicare program (public coverage for the elderly) now spends $15,000 per person enrolled per year. That is nearly twice the national average, and second only to Miami in the United States. In particular, McAllen's costs are almost exactly twice those in the demographically and socio-economically similar Texas border town of El Paso.

McAllen's prominence is recent; in 1992 spending per person enrolled was almost exactly at the national average. None of the standard "justifications" – unhealthy populations, sicker patients, higher quality, malpractice threats – hold any water. Quoting a local surgeon:

"Come on … we all know these arguments are bullshit. There's overutilization here, pure and simple."

In short, the doctors did it.

Gawande's essay has gained tremendous public attention, and possibly some policy relevance, because it was immediately picked up by President Obama, who distributed it to Congressional leaders and said, in effect, "This is what we've got to fix" (Pear, 2008). But McAllen did not catch the President entirely by surprise. Peter Orszag, his budget director and formerly director of the Congressional Budget Office, has repeatedly pointed out that the greatest threat to the fiscal stability of the United States is that posed by rising healthcare costs (Orszag, 2007, 2008a, b). In this context, he has highlighted the central fact of very large regional variations in per-person-enrolled costs in the absence of any evidence of corresponding benefits.

So the picture seems clear. The United States is suffering a steady decline in healthcare sector productivity, including health human resource productivity, of major proportions. Ever-increasing quantities of resources are being poured in, with little or no evidence of corresponding health benefits.

This suggests that policy-makers in any other country with high and growing healthcare expenditures should be looking very skeptically at demands from the industry for ever more resources, and scrutinizing the system carefully to find the places where care is being over-provided: where resources are going in but no benefits are coming out. The key to increased HHRP is simply to reduce the commitment of personnel and other resources to those activities; everything else is just window-dressing. Case closed.

EXCEPT …

It should be ntoed that a major study of the appropriateness of healthcare provision in the United States (McGlynn et al., 2003) found relatively little over-provision of healthcare, but rather very significant *under-provision*. Using the responses from a community tracking survey, linked with medical records, McGlynn et al. were able to identify survey participants' health status over the two years preceding the survey, and the corresponding pattern of use of care. Expert panels, drawing upon clinical evidence, then judged the appropriateness of the care actually given to these individuals, and identified the care they believed should have been given, but was not.

The researchers concluded that study participants were on average receiving little more than half the care appropriate for their conditions:

> "Participants received 54.9 percent … of recommended care. … The deficits we have identified in adherence to recommended processes for basic care pose serious threats to the health of the American public. Strategies to reduce these deficits in care are warranted." p. 2635

Only 11.3% of survey participants were judged to have received care that was "not recommended and was potentially harmful." Far from being over-serviced, Americans would apparently benefit from a great deal *more* care, according to current clinical standards. In a later attempt to measure and classify the forms of waste in the American healthcare system, Bentley et al. (2008) found that only about 2%-3% of American health expenditures could be attributed to items they identified as pure "clinical waste," that is:

> "… spending to produce services that provide marginal or no health benefit over less costly alternatives." p. 644

It is difficult to reconcile these patient-level findings with the aggregate national and regional-level data. Can it really be the case that the country that spends twice as much per capita as the next most expensive in the world falls so far short of providing the range of services that have been demonstrated to be effective in improving health? What might that say about all other countries?

A partial explanation of this paradox, at least for international comparisons, might be that while spending is far above that in any other country, the United States does not provide its citizens with proportionately more healthcare (*on average*). As noted in the sidebar above, a very large share of American health spending is absorbed in the bureaucracy of the private health insurance industry and in the corresponding financial bureaucracies in healthcare institutions. These people provide no healthcare; they are administering a payment system memorably described by Aaron (2003) as:

> "… an administrative monstrosity, a truly bizarre mélange of thousands of payers with payment systems that differ for no socially beneficial reason, as well as staggeringly complex public systems with mindboggling administered prices and other rules expressing distinctions that can only be regarded as weird." p. 801

No other system carries this enormous burden of pure administrative waste. Woolhandler et al. (2003) estimate this waste at $209 bn. U.S. in 1999 or 16.5% of total national health expenditures in that year. Applying the same ratio to officially projected 2009 expenditures of $2,509.5 bn. yields a current estimate of $412.9 bn., or $1,345 for every man, woman and child in the country (as of mid-July, 2009) for the pure administrative waste currently being generated by the peculiarly American forms of healthcare reimbursement.

If this waste could be eliminated, presumably requiring the elimination of the private insurance system itself, and the wasted administrative resources re-deployed to productive uses outside the health sector, the share of American national income spent on healthcare would fall from its currently projected 17.6% in 2009 to 14.7%. The overall productivity of the American health sector would then rise by a remarkable 20%. The implied productivity gain for the whole United States economy is about 3.4% – a very large number indeed. Unproductive activity in healthcare really matters.

Thus, if one abstracts from the costs of the idiosyncratic payment system, actual expenditures for providing healthcare in the United States are very much reduced. However, they remain larger than anywhere else. The paradox of overpayment for underservice remains.

There is an additional possible explanation, noted earlier, and emphasized by Anderson et al. (2003): "It's the prices, stupid!" International comparisons of health expenditures, such as those compiled by the Organization for Economic Cooperation and Development (OECD), adjust national currencies to the U.S. dollar (USD) using either current market exchange rates or purchasing power parities (PPPs). The latter attempt to compare how much the different national currencies will actually buy, in their own countries, from a standard basket of commodities. PPPs are less subject to market fluctuations from capital flows or speculation.

However, both of these conversion factors are defined at the level of the entire economy. In principle, health sector PPPs could be defined specifically for a representative basket of health services and procedures, but this would be very challenging for a number of reasons. In practice, conversions are made using economy-wide PPPs.

It follows that if the prices of healthcare goods and services in a particular country are unusually high relative to those of other non-health commodities – physician visits or surgical procedures, or MRI scans, for example, relative to hamburgers or televisions – unusual, that is, compared to their relationships in other countries, then standard international comparisons of health expenditures converted to USD at economy-wide PPPs will be misleading. They will understate the relative prices of healthcare in the country where healthcare prices are relatively high, and correspondingly overstate the actual quantity of care being provided.

While we have no comprehensive health-sector PPPs for the OECD, we do have quite a bit of partial and service-specific information showing medical prices to be exceptionally high in the United States – e.g. Eisenberg et al. and Antoniou et al. above. A relatively conservative assumption of 50% higher prices in the United States would bring the American share of national income spent on healthcare down from 14.7% (net of administrative waste) to 9.8% – close to the average among high-income countries.

Unlike administrative waste, however, unusually high prices do not imply lower productivity. They say nothing about productivity measured in terms of real outcomes, or even real things done or produced; rather, they generate a transfer of incomes to those who provide the resources used in producing healthcare, from those who pay for it. Resources are being over-reimbursed, not necessarily used inefficiently.

Administrative waste and price inflation thus offer a possible explanation for unusually high healthcare expenditures in the United States, but they do not actually resolve the paradox of apparent under-provision. In the first place, these adjustments still leave that country as a very well-funded system, capable of producing levels of healthcare for its citizens comparable with the rest of the high-income world.

Secondly, and most importantly, they have no bearing on the observed clinical variations within the United States. The variations studies cited above draw data from the public Medicare system for the elderly, and service prices are standardized for regional variations. The regional comparisons of service use are thus not contaminated by variations in either prices or administrative overheads.

An alternative explanation for the McGlynn findings is offered by Dr. Charlyn Black (personal communication). First, although McGlynn et al. suggest that they were looking for both underuse and overuse of services, the indicators that they chose to examine were preponderantly those that would tend to show underuse. Second, it is generally acknowledged that only a small proportion of medical interventions are supported by scientific evidence of effectiveness. A focus on this subset, of the interventions that are "known to work," necessarily excludes a grey area that is likely to be much larger, and in which there is plenty of scope for variations in rates of provision that are driven by the availability of resources, but unrelated to difference in needs or outcomes. This is, as noted above, exactly what Fisher et al. find.

There is a further consideration: the private insurance system not only adds a layer of unproductive expenditure on top of the costs of health services, but it also acts as a powerful and capricious barrier to access. The nearly 50 million Americans who have no coverage may account for a significant share of the under-utilized effective services. Even for those who are insured, the increasing efforts by private (for profit) insurers to limit their own payouts serves as a form of biased "second opinion," a one-way gate. There is no way to compare the (proprietary) standards of "appropriateness" applied to reimbursement decisions by various insurers with the evidence-based standards used by McGlynn et al.

Complaints by American providers about the increasingly intrusive and unreasonable behaviour of insurers have been growing for at least 20 years. One should not, therefore, be surprised if providers respond by increasing the provision of procedures that are easy and lucrative to provide, and difficult for insurers to deny – as in McAllen. It is important not to fall into the trap of assuming that provider responses to economic incentives are even approximately uniform. El Paso differs greatly from McAllen and Minnesota from Miami, for reasons that are not at this point understood (certainly not by economists). However, the tension between provider and insurer incentives, and the strategies and counter-strategies adopted on each side, offer at least a plausible explanation for the paradox of overuse and under-provision.

These considerations do not undermine the value of the McGlynn et al. study; it is powerful (and somewhat disconcerting) in demonstrating the relatively limited uptake of those interventions known to be effective. But this work is not a sound basis for estimating the extent of system overuse, perhaps unsurprising in light of the fact that it was not primarily designed for such detection.

MORE TEA LEAVES – PHYSICIAN AND HOSPITAL ACTIVITY MEASUREMENT

As noted above, the vast majority of productivity studies in healthcare do not address health outcomes; at best they measure activity and outputs. These can give diametrically opposite results depending upon which measures of output are chosen. In Canada, physicians and hospitals provide leading examples.

Studies of physician supply find that while the public (and professional) perception has shifted over the last two decades from "surplus" to "shortage," the actual number of physicians per capita in Canada has not changed for the last 20 years. But has their productivity changed?

Studies of self-reported weekly hours of clinical work of general/family practitioners by Crossley et al. (2008), using survey data from the Canadian Medical Association, have shown that these fell by 15.6% between 1983 and 2003. Watson et al. (2006) also studied this data set, finding an 8.3% decline in hours of work between 1993 and 2003. Preliminary findings from provincial fee payment data in British Columbia (presented at recent meetings of the Canadian Association for Health Services and Policy Research) show average annual days of full-time work to be falling over time for both general practitioners and specialists.

One might interpret this as falling physician productivity, but this would be technically incorrect. The relationship between inputs and outputs, as explained at the outset, applies to outputs per (effort-adjusted) hour, not per capita. If hours of input are falling, output might also be falling but not necessarily output per hour worked.

Perhaps more relevant to the claims of "physician shortage," however, is that output as measured by fee-for-service billings per physician (general practitioners and specialists combined), at standardized fee levels, has *not* been falling over the period during which a shortage is alleged to have emerged. Roughly constant procedural billings, combined with falling hours of work, implies a significant increase in physician productivity – so long as we are able to accept billed services as a measure of physician output that is consistent over time and place.

A further and severe complication is that an increasing proportion of expenditures on physicians' services is not reimbursed on a fee-for-service basis (Alternative Payment Programs or APPs). Salaried and sessional payments to physicians have always made up a small part of the Canadian Medicare system, but in the last decade these have been augmented by various forms of service agreements, regional recruitment and retention payments, payments for on-call services – the list is limited only by the imagination of those negotiating with provincial ministries of health, on behalf of provincial medical associations.

To the extent that such increased reimbursement is not linked to the provision of any specific service, or (like on-call payments) is for services or activities that were previously provided without reimbursement,

these payments simply represent price increases over and above explicit fee schedule changes. They should not enter into any calculation of changing productivity. However, payments that are in fact part of some form of service contract raise the question of the types and amounts of services that are being provided on this basis. Any estimate of changing physician productivity, up or down, must include, either implicitly or explicitly, some estimate of the volumes of services being provided under these arrangements, and the data to support such computations are not generally available.

There is some evidence that the shift to APPs has reduced the average output of services per physician, which might be predicted *a priori* on the basis of elementary economic theory. Why then would provinces accept the expansion of these forms of payment? There are several reasons, but in the background is the long-held concern that fee-for-service distorts patterns of medical practice, encouraging the provision of services with high rates of reimbursement per unit of time required and shortening visit and consultation times. To the extent that this is so, the measurement of service output will show increased "productivity," but the contribution of this activity to health outcomes will be minimal at best.

The dilemma for interpretation is this: if the increased volume of billed activity, combined with services provided through APPs, corresponds to improvement in patient outcomes achieved, then (despite the aging of the physician workforce) physician productivity is increasing quite rapidly. Such a conclusion would undermine claims of present or projected "shortages" based on simple head counts. (As noted above, Watson and McGrail find that Canada enjoys a relatively low rate of mortality relative to other countries in the OECD, despite having one of the lowest doctor-to-population ratios.) And if they do not, then the steady increases in per capita payments to physicians are all, in effect, increases in prices, buying nothing in terms of improved health. The blunt fact is that attempts to estimate the relative balance of productivity increase and price escalation are no more than guesses. However, neither trend appears to offer a compelling case for training more physicians.

The expansion of APPs over the last decade has thus made the estimation/measurement of physician productivity more difficult and less reliable even on an output basis, quite apart from the growing evidence of a disconnect between increasing outputs and improving outcomes. An increasing proportion of payments to physicians is for an unmeasured and unknown volume of services, of unknown effectiveness. There seems to be increasing interest in supplementing APPs with "shadow billing" requirements that physicians reimbursed for service-based APPs report their activities as if they were reimbursed by the provincial fee schedules. It is at least questionable as to whether these reports will yield data comparable with those attached to actual payments, and even if they do, the problems with using billing or payment data to proxy outputs, let alone outcomes, remain firmly in place.

Canada is not alone. A. Scott (2005) reports trends for Australia that have some similarities:

> "In Australia, the quantity of services per doctor has been falling whilst the revenue from fees charged per FTE doctor has been increasing. This suggests that the increase in revenue per doctor is due entirely to increases in real prices. This is against a context of a reduction in the supply of hours by GPs and specialists balanced against weak evidence of an increase in the quality of GP services. Fewer services are being provided and costs are rising with unknown changes in quality." p. 2

Australian doctors, like Canadians, are working fewer hours and collecting increased revenues through higher (inflation-adjusted) prices. Australian doctors have raised fees by increasing the extra-billing of patients; in Canada, doctors have achieved higher prices through the expansion of those APPs that are not linked to services. Canadian physicians are, on average, probably supplying an increasing quantity of services through the combination of fee-for-service payments and service-based APPs, but we do not know how much. In Australia, GPs are supplying fewer services as their hours have fallen, but specialists are maintaining their output levels and charging more, while working fewer hours (Scott, A., 2006).

As A. Scott (2005) notes, however:

> "There are two parts to the measurement of doctor productivity: activities performed and the value of those activities to society, via their impact on health and welfare." *Ibid.*

In neither country do we know much about the latter, but the clinical variations studies are not encouraging.

Another aspect of the measurement of physician productivity that has received surprisingly little attention is the remarkable variation among physicians in apparent measures of productivity. It has long been known that annual physician incomes vary widely within a given specialty and fee regime. To the extent that these reflect differing annual hours of work, they are not, strictly speaking, an indication of differential productivity. But large differences show up among apparently full-time practitioners as well. Studies of average levels of service output, or output per hour, miss the very important and largely unexplained dimension of interpersonal productivity variation. These show up in large standard errors from regression analyses of measures of physician output, variances that tend to outweigh the "explained" variance (Bloor and Maynard, 2001; Crossley et al., 2008).

Unfortunately, increased attention is being given to the construction of more or less elaborate models for projecting physician "requirements." These models are presented with a spurious concreteness, yielding apparently firm answers to important questions. Unfortunately, such models are built on sand. The results they generate are built in from the outset in the strong but highly dubious assumptions that are made about population needs, the effectiveness of physicians' services, other options for delivering those services, and the productivity of physicians themselves. They tell us nothing about HHRP, concealing all the really important questions.

Several such projection exercises carried out in the 1980s had built into their underlying assumptions that all services being provided were "needed" (or else they would not be being provided …) and that service output per physician would and should fall over time because physicians were overworked. The inevitable conclusion was that a physician shortage was about to emerge, even as the physician supply per capita was steadily increasing. This tradition seems recently to have been revived.

Such projection models bypass entirely the question of outcomes, and impose the strong and implausible assumption that current levels and patterns of physician servicing, whatever they may happen to be and however much they may vary across sub-populations, are by definition all and everywhere needed. Unnecessary care simply does not happen. So "needs" expand as the population grows and ages, and may be augmented by projecting trends in particular forms of care use, again without reference to associated outcomes.

Increases in output per physician – for example by expanding the role and availability of various forms of auxiliaries or restructuring patterns of delivery – may be mentioned in passing, but are not quantified for inclusion in the model. In effect, physician productivity is assumed to be static or declining over time. On the other hand, assumptions about average physician time input, which have been trending downward, are quantifiable and can easily be inserted into the projections. The rising trend in output (or at least billable activity) per physician could also be quantified and included in such projections with equal ease, but is somehow overlooked.

These projections tell us nothing about physician productivity in terms of activity and outputs, let alone about health outcomes, because they do not address those issues. Their assumptions are structured such that they can only yield more or less elaborated projections of increasing needs for physicians, which is presumably their purpose.

Such projection models were discredited in Canada by the end of the 1980s. As Lomas, Barer et al. (1985) commented:

> "… funding research that, by its adopted assumptions, will generate answers that are easily determinable without the research is an inefficient and wasteful use of research policy/ planning funds." (p. 113, emphasis in original) (See also Lomas, Stoddart et al., 1985).

But they're back – and not just in Canada. Twenty years later, Maynard (2006) writes:

> "While medical workforce planning is essential, it is generally deeply flawed methodologically and often wrong in the conclusions it reaches. It is usually based on an unquestioning assumption that more healthcare inputs automatically result in more health and that the efficient and unique way to produce improved population health is through increased investments in the health care workforce (Australian Government Productivity Commission 2005). Such an approach should be dealt with sceptically in an industry in which patient outcomes are neither measured nor managed and where a significant proportion of health care in common use has no evidence base (BMJ Publishing Group, 2005)." p. 323

A. Scott (2006), drawing on Maynard (2006), notes that:

> " … medically led workforce planning exercises have been relatively narrow in their scope by focusing solely on recommending changes, most recently large increases, in the number of training places. Such a costly policy does little to address productivity in terms of the inflexibility of training, inappropriate skill mix, or the quality of how doctors practice. Nor does it address inequalities in doctor distribution. Without accompanying structural changes that improve productivity, a policy of increasing the number of doctors is likely to reinforce existing problems and may make them even worse." p. 313

The recurrence of these circular exercises presumably reflects the persistence of their underlying political motivations. In any case, the number of entering medical school places in Canada has risen by about 60% over the last 10 years. Coping with the resulting expansion in physician supply, "reinforce[ing] existing problems," should be sufficient to keep the question of physician productivity off the public agenda for another generation.

HUMAN RESOURCES PRODUCTIVITY IN THE HOSPITAL SECTOR

Physicians generally command centre stage, understandably, given the impact of their decisions across the healthcare system. Payments to physicians, however, account for about twice the volume of resources as absorbed in supporting physicians' services. In this sector, the last 20 years have seen dramatic improvements in productivity of the institutions themselves, and thus implicitly of the personnel who staff them. These major improvements have been largely masked by a focus on outputs rather than outcomes. The budget constraints of the mid-1990s were associated with very large reductions in both separations and patient-days per capita. That such reductions were possible had been known for years, even decades, but the necessary changes in patterns of care were taking place very slowly prior to the stark financial pressures of the 1990s.

The fact that inpatient utilization, and costs, declined so sharply in the 1990s, with no apparent negative (aggregate) effects on patients, is well-known. (The introduction of the Prospective Payment System in the United States, and of the progressively adapted forms of GP fund-holding in the United Kingdom,

had similar effects although at different times.) But these changes have not been widely recognized as dramatic improvements in HHRP. The whole process was, and is, so wrapped up in both large and small "p" political rhetoric – patterns of clinical care underwent forced changes, after all, and jobs and incomes were lost – that the obvious has been obscured. The pharmaceutical industry added further confusion by claiming, implausibly, given the speed of response and its correlation with budgetary changes, that the reductions in hospital use were due to the introduction of new drugs, and thus justified their high prices (Lichtenberg, 1996, 2001).

The significant reductions in requirements for nurses and other hospital personnel resulting from the budget cuts of the mid-1990s have had a reverse echo effect. Reductions in hiring at that time have resulted in shortages 15 years later. The central point, however, is that there has been no (aggregate) evidence of negative consequences for patient outcomes. A good deal of unnecessary inpatient care was forced out of the hospital system by budget cuts. By any reasonable measure, the productivity of that system and the people who staff it has been greatly increased. The HHRP literature, however, seems to be silent on this "elephant in the living room."

Measures of output, rather than outcome, can easily convert these real productivity gains into apparent losses. Staff and other input costs per day tend to rise when lengths of stay fall, as the least intensive and least necessary days are pared away. The large increases in same-day and short-stay surgery represent a major improvement in productivity, but if the same procedure is treated statistically as a different "output" depending upon whether or not it is associated with an inpatient stay, the result is to report the substitution of day surgery for inpatient care as a reduction in hospital output, rather than a reduction in the inputs required to yield a given output – a treated case.

The problem, of course, is that a case is not a case is not a case – these activities are very far from standardized. The same problem arises in measuring physician productivity, but for physicians' services the assumption is generally made, *faute de mieus,* that fee schedules provide an acceptable approximation to the relative amounts of "output" represented by different physician activities. Hospitals present more severe problems; no similar fee schedule exists. Efforts to find ways of standardizing the relative outputs of hospitals with differing case loads have gone on for decades; case mix measurement/adjustment is still an active field of research.

At first blush this (very large) field would not appear to be HHRP-related. However, one need not peel off many layers of the onion to realize that the entire case-mix-adjustment literature is directly relevant to the measurement of hospital productivity, and thus of the productivity of those who work in hospitals. This research seems, in Canada at least, to have largely drifted out of universities, but is actively underway in some provincial ministries of health, and at the Canadian Institute for Health Information.

The Resource Intensity Weights (RIWs) attached to inpatient separations in some provinces in the calculation of hospital budgets, or the Diagnosis-Related Groups (DRGs) used in others, are alternative ways of making more comparable the outputs of hospitals. Such standardized measures of output are essential to comparing the relative efficiency of individual hospitals, using various statistical techniques such as conventional regression analysis and, more recently, data envelopment analysis (DEA), to try to identify more and less efficient hospitals. Since wages and salaries make up the bulk of hospital budgets, relative institutional efficiency would seem to be linked to the productivity of the human resources used.

The observation of very large clinical variations in inpatient use across Canadian regions suggests that there is still considerable scope for further reductions in inappropriate inpatient use and corresponding improvements in HHRP. Hall and Tu (2003), for example, have studied regional variations in rates of hospital admissions across Canada (age-sex adjusted, population 20+) for four major conditions – acute myocardial infarct (AMI), congestive heart failure (CHF), angina, and "chest pain." They found rates ranging from 508.4 per 100,000 people in Vancouver, to 1,929.6 – nearly four times higher – in region 7 of New Brunswick. Rates were markedly lower in metropolitan areas, but there was a pronounced east-west

gradient across metropolitan areas. Rates were about 50 percent higher in central Canada and 75 percent higher in Halifax. The average rate Canada-wide was about double the rate in Vancouver.

A measure developed by McPherson et al. (1982) called the Systematic Component of Variation (SCV) identified the inter-regional variations for the four conditions as ranging from "High" (AMI) through "Very High" (angina, chest pain) to "Very High II" (CHF). Hall and Tu conclude:

> "There is considerable regional variation in the cardiovascular hospitalization rates across the country that may be amenable to further interventional strategies." p. 1123

Such variations obviously raise more detailed questions about comparability of population needs: are people sicker in Halifax or Toronto, or going uncared for in Vancouver? Fisher et al., (2003a, b) did not find poorer outcomes in relatively low-use regions in the United States, where hospitalization rates are on average significantly lower than in Canada; rather, they focused attention on the factors – variations in physician supply and practice styles – that were also the first explanations suggested by Hall and Tu. They also report significant interprovincial variations in average lengths of stay for cardiac cases that tend to correlate with those in admissions, implying even larger geographic variations in patient days per capita.

These observations have received much less attention than they deserve, amid the general clamour about shortages and underfunding. Attempts to make comparisons of relative efficiency among hospitals may have a bearing on HHRP in terms of outputs per unit of labour input, but the much bigger question of productivity with respect to outcomes remains unaddressed. What is the point of attempting to improve the productivity of HHR in the provision of unnecessary hospital care?

There is, nonetheless, a very large literature on HHRP within hospitals, particularly with respect to nursing staff. Most of the studies of nursing productivity turned up by our computerized literature search reported alternative ways of staffing inpatient wards. These fall clearly under the heading of output productivity, i.e. how to maintain a given level of patient care activity (and presumably to maintain patient outcomes) at minimum cost. Within this genre, our computerized search found no studies dealing explicitly with patient outcomes, but as explained in Appendix A, this was a result of the flaw in our search process. The survey of researchers returned a (relatively short) list of studies that reported patient health outcomes associated with differing hospital staffing patterns.

Overall, however, our scoping work suggests that direct (as distinguished from indirect – the variations literature, for example) outcome studies make up a relatively small proportion of the research in this area. This does not mean that these studies are not relevant to questions about HHRP; indeed, if one can find ways to increase more traditional measures of partial factor productivity, and can be confident that patient outcomes are either not affected or (better) actually improve, then such situations do represent unequivocal increases in HHRP, and are worthy of exploration and emulation.

The publication process may, however, be biased, to an unknown degree. We know, from personal contact, of particular innovations in local hospitals that have significantly lowered lengths of stay for particular procedures or improved imaging throughput with little or no requirement for additional resources. But these are not published – the innovators are engaged in providing patient care, not writing for publication. Indeed. there may be a disincentive to publicizing these savings, in case funding agencies react by reducing budgets. This may, to some extent, explain the persistent pattern of productivity-enhancing innovations that are not generalized to other institutions.

HHR PRODUCTIVITY AND HHR PLANNING

There is a major Canadian-based research program on HHR planning, with particular reference to nursing, that specifically recognizes the central significance of assumptions about productivity change, whether explicit or more commonly implicit. Birch, O'Brien-Pallas, Tomblin Murphy and their colleagues combine economic and nursing expertise and have generated a large research output over a number of years (see e.g. O'Brien-Pallas et al. (2007); Tomblin Murphy et al. (2006); Birch et al. (2005); Tomblin Murphy et al. (2003); and O'Brien-Pallas et al. (2001).

Birch et al. (2007), introducing a special supplement of the journal Canadian Public Policy, summarize and apply the conceptual framework that they have developed. There they describe "traditional" approaches to HHR planning in critical terms very similar to those of A. Scott and Maynard, above (and Lomas, Barer et al. and Lomas, Stoddart et al., 20 years earlier).[6] Such "planning" has been essentially an exercise in demographic projection, focusing exclusively on particular provider groups. These projections are based on strong but implicit (and dubious) assumptions about future service productivity (typically unchanging) and future population needs (typically derived from present patterns or trends in age-specific utilization). They demonstrate, with a back-of-the-envelope exercise using data from Atlantic Canada, how dramatically HHR "requirements" projections can fluctuate in response to quite small changes in assumptions about productivity trends.

Birch et al. offer a much more sophisticated four-part framework for projection based on trends in demography, epidemiology, standards of care, and productivity of personnel. This has a powerful logical appeal – how many people (by age and gender), what health problems have they, what level and pattern of services is currently viewed as necessary to address those problems, and how do those service requirements translate into the time and numbers of different types of personnel and associated equipment and supplies?

The authors emphasize that planning, based on assumptions that epidemiology and productivity are "fixed" when they clearly are not, will give inaccurate requirements estimates. Further, as they point out, both patterns and trends in epidemiology and productivity arise in a social, cultural, economic and political context. It would be hard to overemphasize this point, though it is equally hard to quantify these factors in requirements estimates. (This is presumably why several decades of HHR planning, by intelligent people in more or less well-resourced teams, have failed to do so.)

Yet a snake lurks in the planning Eden offered by Birch et al. They identify it in passing, but do not scotch or even pursue it. Its presence shows in their discussion of the American projection of requirements for paediatricians by Shipman et al. (2004). That study estimates that by 2020, the number of children per paediatrician will have fallen by one third. To maintain average workloads (and incomes) they will have to:

> "… provide expanded services to the children currently under their care, expand their patient population to include young adults, and/or compete for a greater share of children currently cared for by non-paediatricians" (Shipman et al., 2004, p.441).

As Birch et al. point out, no consideration is given to the appropriateness of these changed practice patterns in terms of the health of the patients served. Maybe American children (and adults) would be better off if training institutions reduced the production of paediatricians? But that seems not to be an option.

[6] To avoid confusion, note that these authors typically use the acronym "HHRP" to refer to Health Human Resources Planning, not Productivity.

Tomblin Murphy et al. (2006) recognize the same process explicitly:

> "If average needs per person fall over time, there will be too many providers to simply meet needs at current service standards... the quantity of services are expanded independent of changes in needs for services. Moreover, this trend is exacerbated if provider productivity (average services per FTE provider) also increases. Traditionally, the result has been that services per capita increase in order to maintain provider workloads. These higher levels of utilization are then adopted as measures of demand for the next round of HHR planning." p. 2

An increase in the supply of providers can have exactly the same effect as a decline in average needs per person.

"Standards of care" – a key, appropriate and important component of the framework put forward by Birch et al. – are not only variable, but also endogenous. They adjust to accommodate the available supply of personnel because healthcare personnel have sufficient flexibility in what they deem necessary or appropriate. And this, of course, is precisely the interpretation placed by students of the clinical variations literature on the correlation between regional capacity and regional use patterns (see the quote from Fisher (2007) *supra*).

This understanding is not new; Evans and Wolfson (1978) described this process as "moving the target to hit the bullet" – see also Evans (1974). A physician of an earlier day noted that the sickness in a community must be sufficient to employ the physicians in that community.

Planning exercises that derive "standards of care" from current practices will turn out to be chasing their own tails. Lomas, Barer et al. (1985) and Lomas, Stoddart et al. (1985) referred to "supply projections as planning." However, having clearly identified this central problem in their or anyone else's planning framework, Birch et al. apparently go on to treat standards of care as if they were as objective and exogenous as, say, demography.

COMPLEMENTS, SUBSTITUTES, TEAMS – SERVICE ENRICHMENT, YES, PRODUCTIVITY IMPROVEMENT, NO

Moving outside hospitals, there is a long tradition of studies of nurses and other health personnel in expanded clinical roles – generically intermediate-level health practitioners (ILHP) – as substitutes for "peak professionals" such as physicians or dentists. The possibility of substitution is not in itself evidence that ILHPs would represent an improvement in HHRP. The general logic, however, is that such personnel embody a smaller quantity of "human capital" than the more extensively trained peak professional. Training is expensive, both directly and in the form of the "opportunity cost" of the time that trainees must spend out of the workforce. If less-extensively trained personnel can perform the same functions, to equivalent standards of quality – as measured by outcomes – then the higher-level personnel are "overcapitalized."

Substituting the less-highly capitalized personnel to perform those services that they are equally competent to perform represents a smaller commitment of resources – a clear improvement in HHRP – so long as the substitute personnel are as productive, in output per hour terms, as the more highly capitalized professionals and do not require extra complementary inputs such as expensive supervision time. If, however, the "substitute" personnel are employed in different settings and most importantly are paid on different terms – salary versus fee-for-service, for example – one might very well find differences in output per unit time.

There are two waves of this ILHP literature, one from the early 1970s and one arising in the last decade. In both periods, a widespread perception of a physician shortage (rather than a concern for productivity

per se) stimulated a strong interest in substituting nurses with advanced clinical training for general practitioners. Despite a number of successful demonstrations, however, the initiatives of the 1970s came to nothing, and the reasons raise serious warnings about the prospects for similar initiatives today.

The 1970s literature was extensive; by 1981, Jane Record edited an entire small volume reviewing the numerous studies demonstrating that various types of ILHPs could perform many of the functions of general medical practitioners. The conclusions from this research were clear:

> "… new health practitioners (NHPs) appear to perform a large percentage of primary care services at a high level of quality and productivity. Moreover, the gap between the physician/NHP substitution ratio and the NHP/physician cost ratio seems wide enough to assure cost savings when NHPs are used well." (Record et al., 1980, p. 470)

In Canada there was even a randomized controlled trial, the Burlington Randomized Trial of the Nurse Practitioner, that demonstrated the comparable outcomes achieved with these intermediate personnel (Spitzer et al., 1973; Spitzer, 1978). But the rising tide of physicians washed away these possibilities, and in 1984 Walter Spitzer wrote an elegy on "the slow death of a good idea." What happened?

In the 1960s, perceptions of an increasingly severe physician shortage led to a dramatic expansion in the number of medical school training places. However, the projections were largely based on erroneous population projections. The sudden end of the baby boom in the mid-1960s, shortly before the new medical school capacity came on stream, resulted in a 20-year-long unintended expansion in the Canadian physician supply (and associated costs) relative to the population. Large increases in physician supply during the 1970s and 1980s quickly suppressed interest in the nurse practitioner, and the subject lay dormant for about two decades.

In 1990 the Canadian physician-to-population ratio finally topped out and the combination of a growing and aging population, with a stable (per capita) but aging physician supply and a stable number of training places, raised the possibility of a future physician shortage. In this environment, there could have been a very real possibility of substitution of ILHP for physicians (Stoddart and Barer,1992; Barer and Stoddart, 1999).

The authors of a recent Cochrane Collaboration report, (Laurant et al., 2004) reach conclusions re-confirming a generation of research:

> "The findings suggest that appropriately trained nurses can produce as high quality care as primary care doctors and achieve as good health outcomes for patients. … Nurse-doctor substitution has the potential to reduce doctors' workload. However this benefit will not be realised in practise if doctors continue to provide the types of care that have been transferred to nurses. … Nurse-doctor substitution has the potential to reduce the direct costs of care. Cost savings are, however, highly dependent on salary differentials between doctors and nurses … In addition, savings on nurse salaries may be offset by nurses' longer consultation length and increased rate of patient recall relative to doctors …" p. 9

Again, the possibility of substitution has been pre-empted by a sustained and dramatic increase in medical school places, whose graduates are just now beginning to come into practice. Between 1997/98 and 2008/09 the number of first-year students entering Canadian medical colleges rose by 68%, from 1577 (the lowest point since 1970/71) to 2653 (AFMC, 2008). Population growth over the same period has been 11.2%.

The long training and residency periods mean that these increases have only just begun to emerge in the practitioner community, so the rhetoric of "shortage" will no doubt persist for some years yet, as it did in the early 1970s. But the die has already been cast, and the supply of physicians per capita in Canada, which had been stable for the last two decades, is now embarking on another period of expansion that may last for decades. The implications for productivity in this sector are not positive.[7]

Finding the door to clinical practice alongside – or in competition with? – physicians effectively bolted and barred, nurses have turned to an alternative concept of expanded function. Rather than transferring functions from physicians, and thus lowering the size of the required physician stock, they have chosen to carve out their own field of practice and their own clientele. This may involve team practice alongside physicians and others, but providing a different range of services. This is an understandable response to the foreclosure of the possibilities of substitution, but it constitutes an "add-on" of services, not threatening to physicians' professional turf (or incomes). There may well be scope for expanded team practice, particularly in the care of the frail and chronically ill elderly, but we found little or no evidence that this activity would represent an improvement in HHRP.

Maynard (2006) emphasizes the same point with reference to experience in the United Kingdom and the United States. Whatever their capabilities and potential may be, ILHPs in practice do not substitute for physicians. Doctors do not become unemployed. Instead, increased numbers and types of ILPHs result in "service enrichment" – a polite term for higher costs – which may or may not have any benefits in improved outcomes, but almost certainly represents reduced productivity. Productivity improvement means doing more with less; the dominant pattern in healthcare systems has always been doing more with more.

The exception might be a situation of genuine physician shortage, in which expanded numbers and roles of ILHP could be an alternative to expansion of medical school capacity. However, the experience in Canada and elsewhere has been that the expansion of medical schools has foreclosed this possibility.

A very similar lateral shift in ILHP activity shows up even more clearly in dentistry, demonstrating that the process is not a result of the public payment system. In the early 1970s, a perceived shortage of dentists led to a number of experiments with ILHP. In several places, dental hygienists were being trained "in expanded function" to perform much of the work of the general dentist. However, the Saskatchewan Children's Dental Program, initiated in 1974, showed that, for children's dentistry at least, high school graduates with 20 months of training were able to perform the same work, and to quality standards equivalent to those of practicing dentists (as measured in blinded trials). These Dental Nurses (in the Northwest Territories, dental therapists) were able to provide routine dental care for the entire child population of the province. There was no need to build on, and pay for, the skills of the hygienist.

This highly successful program was shut down in the early 1980s by a newly elected government responding to political pressure from dentists. Elsewhere in North America, the profession was successful in suppressing further experimentation with substitutes. This has obvious implications for HHRP, but the veil of silence has been impervious.

Instead of substituting less expensive personnel for dentists, with potentially serious consequences for dentist incomes, there has been an expansion in the provision of traditional dental hygiene services in Canada:

[7] One could plausibly argue that the increased numbers of new physicians will all be needed, and more, if physicians continue to decrease their average workloads. But as noted above, the reported declines in hours of work are not being matched with declines in service output per physician, at least as measured by billings or other forms of reimbursement, nor does this argument factor in the possible impact on physician work decisions, of the great 2008 economic meltdown.

> "The market for the provision of dental hygiene services in Canada generates revenues worth billions of dollars annually. We understand that the delivery of dental hygiene services generates the majority of the revenues in a dental office. We also understand that dental hygiene services generate revenues far in excess of their costs to the dentist." (Scott, S.,2005, 2006)

Regulatory restrictions have required hygienists to work in the offices of dentists, and have thus enabled dentists to appropriate a share of that revenue well above any costs incurred in providing space, equipment, and office services for hygienists. The Canadian Competition Bureau, and dental hygienists themselves, are promoting changes to provincial legislation that would permit "self-initiating" (i.e. free-standing) dental hygiene practice. The Bureau expects (hopes) that this will promote greater competition among hygienists and lower prices for patients. It is unclear, however, that this will improve dental productivity.

The bias toward overcapitalization – excess human capital – is not exclusive to peak professionals such as physicians and dentists. A threat to the HHRP potential of intermediate personnel is the pressure from universities and professional organizations to expand their training requirements to the point that their graduates may become as costly as physicians, particularly if they have shorter "service lives" or years of full-time work per graduate. Do midwives or "physician extenders" under various labels need a full four-year degree in nursing before training in their respective roles? Assurances from university faculties that more of their product – training – is essential may be effective marketing, but they are not outcome-based evidence.

The historical record illustrates several important points about HHRP and personnel substitution. First, research has clearly shown that there is great potential, all else being equal, for productivity improvements from such substitution. Second, all else is never equal. Such productivity improvements represent a threat to the markets of the "over-capitalized" professions and will accordingly be met with hostility and, so far, successful political and organizational resistance. Third, the training and professional organizations themselves can be a threat to improved HHRP insofar as they may dissipate the potential productivity gains through expanding the training requirements and making the substitute personnel unnecessarily expensive. A rigorous focus on outcomes – is the extra training needed, for what, on the basis of what evidence? – can quickly be obscured in the rhetoric of "more is better." And fourth, ascertaining potential gains in HHRP requires a will and an ability to measure outcomes, not just outputs.

LESSONS FOR A SYNTHESIS OF THE EVIDENCE ON HHRP

The current perception of a physician shortage has re-opened the question of physician extenders, not as a way to improve HHRP, but simply as a way to maintain or increase output in the face of a real or impending shortage of physicians. Old questions, apparently settled, are being re-addressed. And one Canadian study of advanced practice nursing, led by Alba DiCenso, has demonstrated with crystal clarity the perhaps most important point to emerge from this scoping exercise.

Despite restricting their scope to the Canadian literature on one relatively small (albeit particularly interesting) sector of healthcare and (primarily) to Canada, they have found over 500 articles. These have taken a nine-person team about a year to read, review and categorize systematically. While the report will be a valuable update of the research in this limited area, there is surely no realistic prospect of carrying out such a comprehensive and in-depth review of all research relevant, directly or indirectly, to the field of HHRP.

If 500 articles required nine people to work for a year, then our estimate of perhaps 12,000 articles (and we do not claim that this estimate represents the entire body of relevant work) would require 24 times that effort (216 person-years). It is not going to happen, nor should it happen, because the average "grade of the ore" found through a comprehensive search is simply not worth that monumental effort

or the associated cost. It would be a very low-productivity use of health services researchers' time. As noted at the outset, questions about HHRP need to be specific and sharply focused if answers are to be feasible. That is Lesson One.

Lesson Two is that virtually all of the literature relevant to HHRP focuses on outputs of procedures or services, not on the health benefits that these outputs may or may not yield. There is general recognition and agreement in principle that it is health outcomes, not healthcare outputs, should be the point and purpose of healthcare, but there is very little reflection of this in the actual studies of HHRP. The great exception to this is the literature on clinical variations; however, there seems to be little, perhaps no, recognition of the implications of that literature for HHRP. A huge conceptual and empirical gap exists between the growing recognition of the scale and significance of clinical variations and the "traditional" HHRP literature largely focused on the production of procedures and services.

Lesson Three is that few of the key decision-makers in the healthcare system *want* to improve HHRP. On the contrary, they resist such improvements energetically and successfully. This should not be surprising: productivity improvement is desired at the economy-wide level, as the basis for improvement in average living standards. For individuals, however, it implies changing habitual patterns of practice, and doing more with less, with a significant threat to jobs and incomes. In any healthcare system, all the incentives faced by the individuals and organizations that are paid to provide care are to do more with more – more outputs, assumed to imply improved outcomes. Productivity improvement faces a classic political dilemma of diffuse benefits and concentrated costs.

That is why the literature shows 40 years of accumulated evidence of the potential productivity gains from substituting less for more extensively trained personnel in clinical applications – and 40 years of complete frustration. New activities and new types of personnel are added on as complementary inputs – service enrichments – but they do not substitute for others, particularly not for the most expensive types of personnel.

It is also why the major improvements in HHRP, in the hospital sector during the Canadian budget squeeze of the 1990s, were so intensely (and effectively) criticized, and have never been recognized as the obvious productivity improvements that they were. When inpatient lengths of stay are shortened, and procedures previously requiring overnight stays are instead performed within a day, with equivalent or even improved outcomes, these are major improvements in HHRP. The fact that the research literature had documented these potential improvements 20 years before, but that very little happened until the budget crisis, underlines the point. There are powerful incentives within healthcare systems that discourage HHRP improvement.

A WAY AHEAD?

While healthcare systems – everywhere, not just in Canada – are characterized by rigidities and resistance to change, it is important to recognize that there are many "points of light" where innovative individuals and institutions have found better ways of providing care. Such innovators are, after all, the reason that we can assert with confidence that day surgery was known to be safe, effective, and efficient 20 years before its major expansion in the 1990s. So, in general terms, the enhancement of HHRP requires the identification of productivity-enhancing innovations, and their generalization system-wide. The former is not so difficult; Priest et al. (2007), for example, provide a collection of success stories of local productivity improvements. The latter generalization is, alas, much more difficult.

The Ottawa Ankle Rules provide a textbook example. An innovative team found, and proved out, a better way of diagnosing ankle injuries in the emergency ward, without routine x-ray. The technique has been universally applauded and has been extended to other injuries. This is precisely the sort of HHRP innovation that the CHSRF, and other funding agencies, should be trying to identify and promote.

A synthesis of research evidence is appropriate only in situations where there is sufficient high-quality evidence to synthesize. In the case of HHRP, we have argued that virtually the entire body of health services research is relevant, directly or (more commonly) indirectly. So there is no question about the sufficiency of evidence. We have also reported that this vast research literature is made up primarily of at best "low grade ore" with a few nuggets. There is some HHRP research of the direct variety, but with very few exceptions, it employs output, rather than outcome, numerators. It is thus of limited value, for the reasons we have articulated at some length in this report. More relevant is the "indirect" research (i.e. where the question that motivated it was not HHRP, but where the results have direct relevance). We have, for the most part, provided the HHRP-relevant synthesis of that work in this report.

Instead, we believe that the CHSRF and its partners should put funding into two types of work. First, they should issue a call for stories of successful HHRP enhancement. The call must make clear that the stories of interest must be about situations in which outputs per unit of input increased *and* there was other evidence of a positive (or at worst neutral) impact on health outcomes. Productivity increase could be indicated either by increased outputs and corresponding identifiable improvements in health outcomes without increased inputs, or by reduced inputs with increases, or at least maintenance of equivalent health outcomes. Those with stories of increased outputs from increased inputs, with no evidence as to health outcomes, need not apply.

The story-tellers should be asked to focus on what it was about the institutional or other circumstances that made it possible to reap these productivity gains – the success factors. The strength of the stories would be further enhanced by some discussion, where relevant, of the extent to which the successful innovations were or were not taken up by other practitioners or institutions, and how this reflected the presence or absence of the relevant success factors. The purpose of this exercise would be both to assemble a collection of productivity-enhancing innovations that have worked in practice, and to find common success factors, so as to assist the generalization of local productivity gains.

Second, CHSRF should issue a call for primary research on HHRP with a focus on the impact on patient outcomes of changes in how health human resources are deployed. The sort of work we envision here could include the evaluation of pilot initiatives or of existing organizational, funding, or other changes, where there are opportunities to examine input (including HHR) use, and resulting outputs and outcomes. These might use either time series or policy trial approaches.

This affords an opportunity for research funding organizations to look forward rather than backward, to become agents of change, facilitating a shift in focus from productivity traditionally measured, to productivity improvements that matter to Canadians.

REFERENCES

Aaron HJ. (2003) The costs of health care administration in the United States and Canada - Questionable answers to a questionable question. N Engl J Med. 2003;349(8):801-803.

Ackman D. (2002) Pay madness at Enron. Forbes.com: <http://www.forbes.com/2002/03/22/0322enronpay_print.html>; March 3, 2002

AFMC - Association of Faculties of Medicine of Canada. (2008) Information about Canadian Faculties of Medicine: 2008/09. Ottawa: <http://www.afmc.ca/pdf/Quotas%2008-09.pdf>; November, 2008

Alter DA, Stukel TA, Newman A. (2008) The relationship between physician supply, cardiovascular health service use and cardiac disease burden in Ontario: Supply-need mismatch. Canadian Journal of Cardiology. 2008;24(3):187-193.

Anderson GF, Reinhardt UE, Hussey PS, Petrosyan V. (2003) It's the prices, stupid: Why the United States is so different from other countries. Health Affairs. May/June 2003;22(3):89-102.

Andrews EL, Baker P. (2009) A.I.G. planning huge bonuses after $170 billion bailout. The New York Times: <http://www.nytimes.com/2009/03/15/business/15AIG.html>; March 14, 2009

Antoniou J, Martineau PA, Filion KB, Haider S, et al. (2004) In-hospital cost of total hip arthroplasty in Canada and the United States. Journal of Bone and Joint Surgery. November, 2004;86-A(11):2435-2439.

Atkinson T. (2005) Atkinson Review: Final Report. London: Palgrave Macmillan; 2005

Australian Government Productivity Commission. (2005) Australia's health workforce. Canberra: Productivity Commission Research Report; 2005

Bachmann LM, Kolb E, Koller MT, Steurer J, ter Riet G. (2003) Accuracy of Ottawa ankle rules to exclude fractures of the ankle and mid-foot: systematic review. Br Med J. Feb. 22, 2003;326:417-423.

Baicker K, Chandra A. (2004) Medicare spending, the physician workforce, and beneficiaries' quality of care. Health Affairs. April 7, 2004;(Web Exclusive):w4.184 - w4.197.

Banks J, Marmot M, Oldfield Z, Smith JP. (2006) Disease and disadvantage in the United States and in England. JAMA. 2006;295(17):2037-2045.

Barer ML, Stoddart GL. (1999) Improving access to needed medical services in rural and remote Canadian communities: Recruitment and retention revisited. Discussion paper prepared for Federal/Provincial/Territorial Advisory Committee on Health Human Resources; June, 1999

Bentley TGK, Effros RM, Palar K, Keeler EB. (2008) Waste in the U.S. health care system: A conceptual framework. Milbank Quarterly. 2008;86(4):629-659.

Birch S, Kephart GI, O'Brien-Pallas L, Tomblin Murphy G. (2005) The Atlantic health human resources planning study. Mississauga, Ontario: Med-Emerg Inc.; 2005

Birch S, Kephart G, Tomblin-Murphy G, O'Brien-Pallas L, Alder R, MacKenzie A. (2007) Human resource planning and the production of health: a needs-based analytical framework. Canadian Public Policy. 2007;33(supp):S1-S16.

Bloor K, Maynard A. (2001) Workforce productivity and incentive structures in the UK National Health Service. Journal of Health Services Research & Policy. 2001;6(2):105-113.

BMJ Publishing Group. (2005) Clinical evidence concise. British Medical Association Publishing: <http://www.clinicalevidence.com/ceweb/products/concise.jsp>; December, 2005

Brehaut JC, Stiell IG, Visentin L, Graham ID. (2005) Clincial decision rules 'in the real world': How a widely disseminated rule is used in everyday practice. Academy of Emergency Medicine. October, 2005;12(10):948-956.

Canadian Institute for Health Information. (2008) Medical imaging in Canada: 2007. Ottawa, Ontario: CIHI; 2008

Crossley TF, Hurley J, Jeon S-H. (2008) Physician labour supply in Canada: a cohort analysis. Health Economics. 2008;Published Online (DOI: 10.1002/hec.1378):1-20.

Cutler DM, McClennan M. (2001) Is technological change in medicine worth it? Health Affairs. 2001;20(5):11-29.

Dawson D, Gravelle H, O'Mahony M, Street A, et al. (2005) Developing new approaches to measuring NHS outputs and productivity. UK Department of Health; September 8, 2005

Donabedian A. (1966) Evaluating the quality of medical care. Milbank Memorial Fund Quarterly. 1966;44:166-203.

Donabedian A. (1989) The end results of health care: Ernest Codman's contribution to quality assessment and beyond. Milbank Quarterly. 1989;67(2):233-256.

Eisenberg MJ, Filion KB, Azoulay A, Brox AC, Haider S, Pilote L. (2005) Outcomes and cost of coronary artery bypass graft surgery in the United States and Canada. Archives of Internal Medicine. 2005;165:1506-1513.

Evans RG. (1974) Supplier-induced demand: Some empirical evidence and implications. In: Perlman M, editor. The Economics of Health and Medical Care (Proceedings of a conference held by the International Economic Association in Tokyo). London: MacMillan; 1974. p. 162-173.

Evans RG. (1995) Healthy populations or healthy institutions? The dilemma of health care management. Journal of Health Administration Education. Summer, 1995;13(3):453-472.

Evans RG. (2007) Extravagant Americans, healthier Canadians: The bottom line in North American health care. In: Thomas DM, Torrey BB, editors. Canada and the United States: Differences that Count. Peterborough, Ont.: Broadview Press; 2007. p. 135-164.

Evans RG. (2008) Political arithmetick: physician productivity in concept and measurement (Draft Discussion Paper). Vancouver: UBC Centre for Health Services and Policy Research; November, 2008

Evans RG, Wolfson AD. (1978) Moving the target to hit the bullet: Generation of utilization by physicians in Canada. Stanford: Paper prepared for the NBER Conference on the Economics of Physician and Patient Behaviour; January 27-8, 1978:Unpublished.

Feldstein M. (1974) The quality of hospital services: An analysis of geographic variation and intertemporal change. In: Perlman M, editor. The Economics of Health and Medical Care (Proceedings of a conference held by the International Economic Association in Tokyo). London: MacMillan; 1974. p. 402-419.

Fisher ES. (2007) Robert and Alma Moreton Lecture: Pay-for-performance: more than rearranging the deck chairs? Center for the Evaluative Clinical Sciences, Dartmouth Medical School; May 24, 2007: PowerPoint Presentation.

Fisher ES, Wennberg DE, Stukel TA, Gottlieb DJ. (2004) Variations in the longitudinal efficiency of academic medical centers. Health Affairs. October 7, 2004;(Web Exclusive):var19 - var32.

Fisher ES, Wennberg DE, Stukel TA, Gottlieb DJ, Lucas FL, Pinder EL. (2003a) The implications of regional variations in medicare spending. Part 1: the content, quality, and accessibility of care. Ann Intern Med. 2003;138(4):273-287.

Fisher ES, Wennberg DE, Stukel TA, Gottlieb DJ, Lucas FL, Pinder EL. (2003b) The implicaitons of regional variations in medicare spending. Part 2: health outcomes and satisfaction with care. Ann Intern Med. 2003;138(4):288-298.

Gawande A. (2009) The cost conundrum: What a Texas town can teach us about health care. The New Yorker. June 1, 2009

Geyman J. (2008) Do not resuscitate: Why the health insurance industry is dying and how we must replace it. Common Courage Press; 2008:253.

Goodman DC, Fisher ES, Little GA, Stukel TA, Chang C-H, Schoendorf KS. (2002) The relation between the availability of neonatal intensive care and neonatal mortality. N Engl J Med. 2002;346(20): 1538-1544.

Goodman DC, Fisher ES. (2008) Physician workforce crisis? Wrong diagnosis, wrong prescription. N Engl J Med. April 17, 2008;358(16):1658-1661.

Goodman DC, Grumback K. (2008) Does having more physicians lead to better health system performance? JAMA. January 23, 2008;299(3):335-337.

Hall RE, Tu JV. (2003) Hospitalization rates and length of stay for cardiovascular conditions in Canada, 1994 to 1999. Canadian Journal of Cardiology. 2003;19(10):1123-1131.

Krugman P, Wells R. (2006) The health care crisis and what to do about it. New York: The NY Metro Chapter of Physicians for a National Health Program, The Community Church of NY Action for Justice Committee, The New York Review of Books; May 30, 2006:A public forum at the Community Church of New York.

Lasser KE, Himmelstein DU, Woolhandler S. (2006) Access to care, health status, and health disparities in the United States and Canada: Results of a cross-national population-based survey. American Journal of Public Health. July, 2006;96(7):1300-1307.

Laurant M, Reeves D, Hermens R, Braspenning J, Grol R, Sibbald B. (2004) Substitution of doctors by nurses in primary care. Cochrane Database of Systematic Reviews. 2004;Issue 4:Art. No.: CD001271. DOI: 10.1002/14651858.CD001271.pub2.

Lichtenberg FR. (1996) Do (more and better) drugs keep people out of hospitals? American Economic Review. 1996;86(2):384.

Lichtenberg FR. (2001) Are the benefits of newer drugs worth their cost? Evidence from the 1996 MEPS. Health Affairs. 2001;20(5):241-251.

Lomas J, Barer ML, Stoddart GL. (1985) Physician manpower planning: Lessons from the Macdonald Report. Toronto: Ontario Economic Council (OEC Discussion Paper Series); 1985:viii, 128.

Lomas J, Stoddart GL, Barer ML. (1985) Supply projections as planning: A critical review of forecasting net physician requirements in Canada. Social Science & Medicine. 1985;20(4):411-424.

Madrian BC. (1994) Employment-based health insurance and job mobility: Is there evidence of job-lock? Quarterly Journal of Economics. 1994;109(1):27-54.

Maynard A. (2006) Medical workforce planning: some forecasting challenges. The Australian Economic Review. 2006;39(3):323-329.

McGlynn EA, Asch SM, Adams J, Keesey J, et al. (2003) The quality of health care delivered to adults in the United States. N Engl J Med. June 26, 2003;348(26):2635-2645.

McPherson K, Wennberg J, Hovind O, Clifford P. (1982) Small-area variations in the use of common surgical procedures: An international comparison of New England, England and Norway. N Engl J Med. 1982;307:1310-1314.

Nolte E, McKee CM. (2008) Measuring the health of nations: Updating an earlier analysis. Health Affairs. 2008;27(1):58-71.

O'Brien-Pallas L, Tomblin Murphy G, Birch S, Baumann A. (2001) Framework for analyzing health human resources (Figure 1). In: Future development of information to support the management of nursing resources: Recommendations. Ottawa: CIHI; 2001. p. 6.

O'Brien-Pallas L, Tomblin Murphy G, Birch S, Kephart G, et al. (2007) Health human resource modelling: challenging the past, creating the future. Ottawa: CHSRF/FCRSS; October 31, 2007

Organization for Economic Co-operation and Development. (2009) OECD Health Data 2009. Paris: OECD; July 1, 2009

Orszag PR. (2008a) Geographic variation in health care spending. Washington, D.C.: Congressional Budget Office - Congress of the United States; February, 2008

Orszag PR. (2008b) The overuse, underuse, and misuse of health care: Testimony before the Committee on Finance, United States Senate. Washington, D.C.: Congressional Budget Office - Congress of the United States; July 17, 2008

Orszag PR, Ellis P. (2007) The challenge of rising health care costs - A view from the Congressional Budget Office. N Engl J Med. November 1, 2007;357(18):1793-1795.

Pear R. (2009) Health care spending disparities stir a fight. New York Times: <http://www.nytimes.com/2009/06/09/us/politics/09health.html>; June 8, 2009

Perry JJ, Stiell IG. (2006) Impact of clinical decision rules on clinical care of traumatic injuries to the foot and ankle, knee, cervical spine, and head. Injury. 2006;37:1157-1165.

Record JC. (1981) Staffing primary care in 1990: Physician replacement and cost savings. New York, NY: Springer Publishing Co.; 1981

Priest A, Rachlis M, Cohen M. (2007) Why wait? Public solutions to cure surgical waitlists. A submission to the BC Government's Conversation on Health. Vancouver: Canadian Centre for Policy Alternatives; May, 2007: 40p.

Record JC, McCally M, Schweitzer SO, Blomquist RM, Berger BD. (1980) New health professions after a decade and a half: Delegation, productivity and costs in primary care. Journal of Health Politics, Policy and Law. Fall, 1980;5(3):470-497.

Schoen C, Doty MM, Collins SR, Holmgren AL. (2005) Insured but not protected: how many adults are underinsured? Health Affairs. June 14, 2005;(Web Exclusive):w5289-w5302.

Scott A. (2005) The productivity of doctors in Australia: the 'flat of the curve' and beyond? Melbourne: The University of Melbourne, Melbourne Institute Working Paper Series: Working Paper No. 19.05; 2005

Scott A. (2006) The productivity of the health workforce. The Australian Economic Review. 2006;39(3):312-317.

Scott S. (2005) Dental Hygienists' Act - An act respecting the regulation of the profession of dental hygiene. Ottawa: Competition Bureau, Government of Canada; December 22, 2005: <http://www.competitionbureau.gc.ca/eic/site/cb.

Scott S. (2006) Dental Hygiene Act, 2005. Ottawa: Competition Bureau, Government of Canada; January 27, 2006: <http://www.competitionbureau.gc.ca/eic/site/cb.

Sharpe A, Bradley C. (2008) The measurement of output and productivity in the health care sector in Canada: an overview. Ottawa: Centre for the Study of Living Standards; 2008

Shipman S, Lurie J, Goodman D. (2003) The general pediatrician: projecting future workforce supply and requirements. Pediatrics. 2004;113:435-442.

Skinner JS, Staiger DO, Fisher ES. (2006) Is technological change in medicine always worth it? The case of acute myocardial infarcation. Health Affairs. Feb 7, 2006;(Web Exclusive):w34-w47.

Spitzer WO, Kergin DJ, Yoshida MA, et al. (1973) Nurse practitioners in primary health care: III. The Southern Ontario randomized trial. CMAJ. 1973;108:1005-1016.

Spitzer WO. (1984) The nurse practitioner revisited: slow death of a good idea. N Engl J Med. April 19, 1984;310(16):1049-1051.

Spitzer WO. (1978) Evidence that justifies the introduction of new health professionals. In: Slayton P, Trebilcock MJ, editors. The professions and public policy. Toronto: University of Toronto Press; 1978. p. 211-236.

Stiell IG, Greenberg RD, McKnight RD, Nair RC, McDowell I, Worthington JR. (1992) A study to develop clinical decision rules for the use of radiography in acute ankle injuries. Annals of emergency medicine. 1992;21:384-390.

Stiell IG, McKnight RD, Greenberg RD, McDowell I, et al. (1994) Implemenation of the Ottawa ankle rules. Journal of the American Medical Association. 1994;271:827-832.

Stockwell H, Vayda E. (1979) Variaions in surgery in Ontario. Med Care. 1979;17:390-396.

Stoddart GL, Barer ML. (1992) Toward integrated medical resource policies for Canada: 11. Improving effectiveness and efficiency. Canadian Medical Association Journal. Dec. 1, 1992;147(11):1653-1660.

Tomblin Murphy G, Birch S, Wang S, O'Brien-Pallas L, Alder R, Betts H. (2006) Report to the Canadian Nurses Association (Canadian Nurse Practitioner Initiative): development of health human resources simulation model for nurse practitioners in primary health care. Mississauga, Ontario: Med-Emerg Inc.; March, 2006

Tomblin Murphy G, O'Brien-Pallas L, Alksnis C, Birch S, et al. (2003) Health human resources planning: an examination of relationships among nursing service utilization, an estimate of population health and overall health status outcomes in the province of Ontario. Ottawa: CHSRF/FCRSS; November, 2003

Tu JV, Ghali WA, Pilote L, Brien S, editors. (2006) Canadian cardiovasular atlas. Toronto: Canadian Cardiac Outcomes Research Team, Institute for Clinical Evaluative Sciences (ICES); 2006

United Nations Statistics Division. (1993) UN system of national accounts 1993. <http://unstats.un.org/unsd/sna1993/introduction.asp> ; 1993

Vayda E, Morrison M, Anderson GD. (1976) Surgical rates in the Canadian provinces, 1968-1972. Canadian Journal of Surgery. 1976;19:235-242.

Vayda E. (1973) A comparison of surgical rates in Canada and in England and Wales. N Engl J Med. December, 6 1973;289(23):1224-1229.

Wagstaff A. (2009) Social health insurance vs. tax-financed health systems - Evidence from the OECD. Washington, D.C.: World Bank: Policy Research Working Paper #4821; January, 2009:36p.

Watson DE, McGrail KM. (2009) More doctors or better care? Healthcare Policy. August, 2009;5(1):26-31.

Watson DE, Slade S, Buske L, Tepper J. (2006) Intergenerational differences in workloads among primary care physicians: A ten-year, population-based study. Health Affairs. November/December 2006;25(6):1620-1628.

Wennberg DE. (2005) Variation in use of medicare services among regions and selected academic medical centers: Is more better? (Commonwealth Fund Pub. no. 874). Commonwealth Fund; December, 2005:43p.

Wennberg DE, Skinner JS, Fisher ES. (2002) Geography and the debate over medicare reform. Health Affairs. Feb. 13, 2002;(Web Exclusive):w96-w114.

Woolhandler S, Campbell T, Himmelstein DU. (2003) Costs of health care administration in the United States and Canada. N Engl J Med. Aug. 21, 2003;349(8):768-775.