Political Arithmetick:
Physician Productivity in Concept and Measurement

November 2008
Draft Discussion Paper

Robert G Evans PhD
Department of Economics and
Centre for Health Services and Policy Research, UBC
About CHSPR

The Centre for Health Services and Policy Research (CHSPR) is an independent research centre based at the University of British Columbia. CHSPR’s mission is to advance scientific enquiry into issues of health in population groups, and ways in which health services can best be organized, funded and delivered. Our researchers carry out a diverse program of applied health services and population health research under this agenda. The Centre’s work is:

- Independent
- Population based
- Policy relevant
- Interdisciplinary
- Privacy sensitive

CHSPR aims to contribute to the improvement of population health by ensuring our research is relevant to contemporary health policy concerns and by working closely with decision makers to actively translate research findings into policy options. Our researchers are active participants in many policy-making forums and provide advice and assistance to both government and non-government organizations in British Columbia (BC), Canada and abroad.

Funding and Support

CHSPR receives core funding from the BC Ministry of Health, and ongoing support from the University of British Columbia. This enables the Centre to focus on research that has a direct role in informing policy and health reform, and facilitates CHSPR’s continuing development of the BC Linked Health Database.

Our researchers are also funded by competitive external grants from provincial, national and international funding agencies. They include the Canadian Health Services Research Foundation, the Canadian Institutes of Health Research, the Commonwealth Fund, Health Canada, the Michael Smith Foundation for Health Research, and WorkSafeBC.

Data Services: BC Linked Health Data Base

Much of CHSPR’s research is made possible through the BC Linked Health Database, a valuable resource of data relating to the encounters of BC residents with various health care and other systems in the province. These data are used in a de-identified form for applied health services and population health research deemed to be in the public interest.

CHSPR has developed strict policies and procedures to protect the confidentiality and security of these data holdings and fully complies with all legislative acts governing the protection and use of sensitive information. CHSPR has over 30 years of experience in handling data from the BC Ministry of Health and other professional bodies, and acts as the access point for researchers wishing to use these data for research in the public interest.
Foreword

This paper was written for the Fourth Annual Medical Workforce Conference, San Francisco, November 4-7, 1999. The organizers’ intent was to have the conference papers published, and to have the discussants for the papers in each session contribute a synthesis paper for that session. The papers synthesized here are:

Harding, John, and Warwick Conn, “Workforce Productivity in the Australian Medical workforce”

Watanabe, Mamoru, Lynda Buske and Jill Strachan, “Canadian Physician Workforce Productivity”


Unfortunately the organizers’ plans for publication did not materialize. This paper reached the “antepenultimate” stage of being circulated to the session participants for comment and possible revisions, after which references were to have been added. But the overall project was abandoned before any comments were received, and the paper was never completed for publication.

The broader issue of physician productivity has, however, re-emerged in Canada. I believe that the paper offers a useful analytic framework for addressing that topic along with fairly detailed illustrative examples of its application to the descriptive material, institutional and statistical, provided in the session papers. Much has changed since; no attempt has been made to up-date those papers. But the process of applying the analytic framework to the world “as it then was” (said to be) provides, I think, a clear guide for a similar application to present circumstances.

It should be emphasized that the authors of the original papers have not reviewed this draft; responsibility for errors, omissions, and mis-representations rests with me.
"Ye shall no more give the people straw to make brick, as heretofore; let them go and gather straw for themselves.

And the tale of the bricks, which they did make heretofore, ye shall lay upon them; ye shall not diminish ought thereof; for they be idle, therefore they cry, saying Let us go and sacrifice to our God.

Let there be more work laid upon the men, that they may labour therein; and let them not regard vain words." Exodus 5:7-9.

Brick production required three inputs -- mud, straw, and Israelite-hours. “You cannot make bricks without straw.”¹ When Pharaoh refused to provide it, the Israelites had to glean the necessary fibre wherever they could. (Mud was a “free good”.) Yet Pharaoh demanded the full tale of bricks, so the Israelites had to work longer hours. They were not pleased.

The Exodus story illustrates four points about productivity measurement that are relevant to the somewhat more complex world of physicians’ services and health care.

First, all inputs must be identified and appropriately measured. “Bricks per Israelite” did not change after Pharaoh’s order, but “Israelites” were not an input. Brick production used up Israelite-hours (standardized for effort and skill), not Israelites themselves. And productivity measured as bricks per Israelite-hour fell, even though the “tale of bricks” -- output per Israelite-day or -week -- did not. If the inputs are specified incorrectly, productivity measures will give erroneous results.

Second, straw and Israelite-hours were substitutes in production; less of one could be made up with more of the other to keep total output constant. Any measure of output relative to the amount of a particular input – such as bricks per Israelite-hour -- will be a function of the available supplies of other inputs.² It will in general be misleading to measure the “productivity” of any one input – and to estimate the amounts needed -- in isolation from information on the supply and substitutability of other inputs.

Third, there is an important difference between average and marginal productivity. When straw was being provided, one could have calculated the ratio (by weight) of brick output to straw input and called that the average productivity of straw. But at that point (presuming that Pharaoh’s overseers were providing all that was asked for) additional straw would have added nothing to brick output. [Too much straw, and the brick falls apart.] The marginal productivity of straw was zero. But after the change in policy, the marginal productivity of straw went up sharply. Holding Israelite-hours constant, an increase in straw supply would have added significantly to total brick output.

But that leads to the fourth and probably the most important lesson from the Exodus story – it is not about productivity in the Egyptian brick industry. Nobody cared. Pharaoh and the Israelites had more important concerns on their minds; and the observed changes in productivity were simply side-effects that emerged from the pursuit of other policy objectives. If there was an “efficiency

¹ You can, but they break. Mud brick is a composite material like fibreglass, in which the straw prevents the propagation of cracks in the brittle matrix. “Fibreglass” structures made only out of resin, with no embedded glass fibres, would simply crack and crumble, like the bricks.

² Even if the technology is “Leontief” output per unit of input X will still depend upon the supply of input Y, though the function will have a rather simple form.
expert” among Pharaoh’s overseers – and there may well have been -- he must have been a very frustrated man.³

Pharaoh had a serious political problem. His country was economically dependent on a large number of gast-arbeiter whose numbers were growing rapidly such that they might soon threaten his and his people’s control of the land of Egypt. He could not let them go,⁴ and he had tried and failed to limit their numbers. Now outside agitators – Moses and Aaron – were stirring up the collective ethnic consciousness of Israel. Withholding straw would at least keep them busy, with less time to get together, and might also discredit the agitators. Israel, on the other hand, were groaning under the increasingly harsh Egyptian oppression, and trying to decide how far to go with a major religious revival.

Economic efficiency entered no one’s calculations. Pharaoh did not pay for the Israelites’ time; and the Israelites certainly had no economic interest in brick production. They were obviously unhappy to have more of their labour time appropriated, and a disgruntled work-force might require more (costly) oversight. But any such cost was lost in the much larger conflict of political, ethnic and religious interests.⁵

This fourth observation may explain the curious opening to the paper by Maynard and Bloor. They propose: “… to present an initial exploration of issues associated with the productivity of the physician workforce …” in the U.K. Initial? The National Health Service celebrated its fiftieth anniversary in 1998. Throughout that half century, the level and extent of services provided by the NHS have been matters of intense political debate and controversy. Yet after all that time, we are offered an initial exploration of physician productivity issues. Could it really be true, that amid the endless rhetoric of “underfunding,” waiting lists, and rationing, this central question simply slipped everyone’s mind? Unlikely. Yet it has clearly not been adequately addressed either. Such curious behaviour, like that of the dog in the night-time, is highly suggestive. Nor is it unique to the U.K.

A private, for-profit firm in the business of producing medical services – or bricks – would worry a lot about productivity. Management, observing that physician-hours are a relatively expensive input, might take a very active interest in economizing on their use, and finding less expensive substitutes. Whether or not they succeeded, it is hard to believe that the issue would remain unexplored for fifty years.

There are of course a number of reasons, many of them very good ones, why we do not entrust the production of physicians’ services to for-profit corporations.⁶ But the NHS, though not a for-profit corporation, is a public agency responsible for meeting the health care needs of the British people (as far as possible) within a given budget. One might have thought that such an agency would be very

³ The author is aware of the dubious historicity of the Exodus account. Like the rest of the Bible, its value is didactic.
⁴ Letting them go probably would have been the best policy, but the Lord had hardened Pharaoh’s heart. He was being set up for a demonstration to Israel that the God of their fathers was the One True God. An economist might say that he was being led, as if by an invisible hand, to encompass an end that was no part of his intention.
⁵ Translating into Economese, Pharaoh attached a negative value to the input, Israelite-hours. That commodity was a “bad”, not a good, and was not freely disposable, so using more of it up in brick production made him better off. And the negative price of this input had become larger, absolutely, since Moses and Aaron had arrived.
⁶ For the last decade or so, Americans seem to have been determined to re-discover these reasons for themselves. The wise learn from the experience of others.
concerned about the productivity of one of its most expensive inputs, keeping good statistics on the subject, and constantly striving for improvement. Apparently not.

Nor does the improvement of physician productivity, or even its measurement, appear to be a central concern in any of the other three countries represented at the Trilateral Conference. Such consistency should remind us of the paraphrase of Hegel by the political scientist Theodore Marmor: “Nothing that is regular, is stupid.” Or the litigator’s advice: “Never ask a question, if you do not already know the answer.” If we find physician productivity badly measured, poorly analysed, and largely ignored by policy in all modern health care systems, it is probably not because the key participants were stupid, or forgot. Like Pharaoh and the Israelites, they have had other pressing concerns, and often conflicting objectives, that took precedence over productivity, or simply made the issue politically too hot to handle.

Productivity, after all, is as much a political as an economic issue, if not more so. There are good reasons why workers hate the “efficiency expert”. “In the long run”, and “on average” societies become wealthier only through increasing productivity per worker. But the average may lump my loss with your gain, and in the long run we are all dead.

The ambiguous attitudes toward physician productivity come into sharp focus in Canadian health policy. Each of the Canadian provinces operates a public health insurance system that reimburses physicians by fees per item of service, according to a uniform fee schedule negotiated periodically with the provincial medical association, and without payments by patients.

Average service volumes per physician, or at least billings adjusted for fee change, have risen steadily over time. This could be interpreted -- and celebrated -- as increasing productivity, more “output” per physician. But provincial governments see it quite differently. They see an uncontrolled and growing expenditure component in their budgets. Worse, this “productivity” component tends to

---

7 There is a nice example of the political use of “productivity” claims that goes back to a hypothesis of William Baumol in the early 1960s. Baumol postulated that over the long term, productivity must increase more slowly in the service industries than in the goods producing industries. It would follow, then, that if prices and incomes are set in perfectly competitive markets, where all firms minimize costs, service prices would have to rise over time relative to goods prices, in order to permit service producers to offer wages competitive with those in the goods industries. In particular, the price of health care services should rise relative to the general price level. Medical “inflation” is not a problem, but a natural side effect of rapid productivity in the goods-producing industries.

The logic was valid; both underlying assumptions were wrong. Productivity has in fact risen very rapidly in some sectors of health care, less so in others, as it has in the goods industries. There is no evidence of some fundamental technological bias one way or another. Moreover the behaviour of medical prices do not even remotely approximate those that might emerge in a competitive market – and why should they? Dramatic productivity increases for specific types of surgery have not, even in the U.S., resulted in a fall in their relative prices. That has been enforced by legislative action -- the Medicare RBRVS. And the price of health care, relative to the general price level, is much higher in the supposed “competitive” U.S. than in any other country in the OECD.

Yet Baumol’s story has continued to surface from time to time in the U.S. literature in particular, because it provides a pseudo-explanation for the exceptionally rapid rates of price inflation in that country. It can be used by the beneficiaries of that inflation to discourage public regulation, and has thus become an intellectual coprolith. Ironically Baumol was influenced by consideration of the performing arts, and the observation that it is difficult to improve the productivity of a string quartet so as to use only three people. As he was writing, solid-state technology was being applied to mass-market sound reproduction and colour television was being introduced. Stereo LPs have since given way to the CD, DVDs are reaching the mass market, and HD television is on its way. Wall-sized 3-D screens are a little further out. It is quite important to get one’s output measures right, when discussing productivity!
rise more rapidly, the more unsatisfied physicians are with their fee increases. (Precisely the same process has been observed in the U.S. Medicare program, where the rule of thumb is that 50% (QZ CHK REF) of any reduction in average fees will be off-set by a volume response.) Total expenditures are the product of prices and quantities, $P \times Q$, and attempts to limit $P$ induce off-setting increases in $Q$.

Accordingly provincial governments have attempted to impose caps on total payments, such that any increases in $Q$ trigger “claw-backs” or off-setting reductions in $P$ (QZ REF). In effect they are trying to discourage, or at least penalize, these increases in “productivity”. (U.S. Medicare imposes Volume Performance Standards for the same reason.)

Meanwhile, in each of the countries represented at the Trilateral Conference there has been a rapid increase in the proportion of women physicians, and a corresponding universal observation that the “productivity” (variously measured) of women is less, on average, than that of men. It follows that an increasingly “feminized” physician work-force must also be expanded to maintain current levels of service. Medical school classes must be increased, stat. This line of argument, which is actually quite old, is currently being used by medical school deans in Canada, and the Association of Canadian Medical Colleges (ACMC), as part of a very aggressive campaign to convince Canadians of a looming physician shortage, and to capture for their schools a share of the emerging surplus in public budgets.

The “shortage” claim is controversial; the policy observations are not. Provincial reimbursement policies are designed to try to hold down increases in the rate of billings per physician, in order to control costs. Medical school representatives are using a projected fall in that rate, inter alia, as an argument for increasing the supply of physicians – and thereby, billings. The two policies are directly contradictory. Yet the two debates, over expenditure control and physician supply, remain in rhetorically splendid isolation, conversations of the deaf (primarily in the newspapers).

Since the parties involved are (for the most part) neither ignorant nor stupid, it might be worthwhile to consider why.

Figure 1 offers as an organizing framework for this question, a set of linked concepts that underlie productivity discussions. “Capacity” refers to the stock of personnel and other assets, forms of capital, from which the inputs to production can be drawn. These capital assets may be physical (buildings and equipment), human (training and skill embodied in different types of personnel, or disembodied “know-how”. All are “produced means of production”. Schematically, we may think of a vector “$N$”, a string of values of which the $i^{th}$ element, $n_i$, represents the quantity available of a particular asset of type $i$ – hospital beds, for example, or ophthalmologists, or lithotripters.
Figure 1

The Health Care Production Chain: From Resources to Health

CAPACITY
Numbers and types of personnel with embodied skills and knowledge, physical facilities and equipment, “Know-how” -- the vector “N”.

\[
\text{\textbackslash/}
\]

INPUTS
Person-hours, by skill and energy level, services of various forms of capital, all devoted to the production of various forms of health care -- the vector “Z”.

\[
\text{\textbackslash/}
\]

OUTPUTS
Quantities produced and used of all the different types of goods and services comprised in the term “health care” – the vector “Q”.

\[
\text{\textbackslash/}
\]

OUTCOMES
(Positive) changes in health status, as a result of use of the quantities in the vector Q.

The objective may be “Health” for the population, or it may be some more general measure of happiness or satisfaction. The latter may depend more upon perceptions, including those generated by access to and use of care, than on actual outcomes. [What then?]
Like Israelites, however, these assets are not themselves used up in production – at least not all at once. But they yield a flow of services, inputs to production, that can be measured (for physical and human capital) in terms of rates of use – Israelite-hours, for example, or hours of OR time. These are combined with other resource inputs – straw and mud, or kWh, or drugs and supplies. The string \( Z \) is the vector of amounts of all of these different inputs used up in the production of health care goods and services, and element \( z_i \) might be, for example, nursing hours.

The services themselves, represented by the vector \( Q \), may be variously grouped depending upon the organization of a particular health care system. In the NHS, Finished Consultant Episodes (FCEs) would now represent one element of \( Q \), or this could be subdivided into FCEs of different specialty types. In fee for service systems, the elements of \( Q \) tend to be defined by the items in the fee schedule – office visits, consultations, specific diagnostic or surgical procedures. But the elements of \( Q \) also include services of other practitioners (working independently), amounts of different types of drugs used, and services provided by institutions themselves such as patient days or separations from hospitals.

Finally the elements of \( Q \) derive their value from their contribution to the health of the patients to whom they are provided (and possibly others, as in the case of immunization). The vector \( H \) represents the health status of each of the members of the served population, and the presumption is that the mix and volume of the elements of \( Q \) is guided by their expected (positive) impact on \( H \).

Expectations are sometimes disappointed, but if on average a good or service can reasonably be expected to improve the health of those to whom it is offered, then it should be provided. And if there are identifiable people who could benefit, but are not (“unmet need”), then the service should be expanded. On the other hand, if overall and on average, a service does no discernable good, or \textit{a fortiori} does more harm than good, it should be provided more selectively or not at all.

This framework permits us to pin-point the sources of the Canadian policy contradictions, and perhaps to understand the reluctance of participants to resolve them. Implicit in the policy of global caps is the assumption that some increases in billings per physician – especially those that follow limitations on fees – correspond to increases in service volumes that either do no good for patients (\( Q \) up, \( H \) constant, zero \textit{marginal} productivity of health), or do not occur at all. Apparent increases in “productivity” are then simply disguised price increases. Absent those assumptions, and the

---

8 They, and we, depreciate over time, but in complex ways. Human capital, unlike physical, depreciates from lack of use, not use, but also depreciates purely from age. All forms may become obsolete, but “know-how” does not depreciate at all – it \textit{only} becomes obsolete.

9 Patient days are the more appropriate measure of output in long-term care, while acute care output is better measured as separations – but there are well-known ambiguities here.

10 This normative principle underlies the universal health care systems in Australia, Canada, and the U.K. The U/S/ is profoundly ambiguous on this point; most Americans claim to agree with it, but reject the policies necessary to give it effect.

11 What is discernible? Welch QZ has labelled as “epsilon services” those that do have some (expected) health benefit, but very little, and at great cost. Should those very small benefits be treated as zero, or as “needs” to be met? The answer tends to depend upon whether one is paying, or being paid, for these services – but their numbers may be infinite.

12 This is not necessarily fraud. The definitional boundaries between different fee schedule items are inevitably fuzzy, and the billings associated with a given pattern of work can be varied considerably within perfectly legal limits by “creative” billing. And different practice styles and patterns can yield very different billing rates, for the same input of time and effort.
provincial policy amounts to refusing to pay for needed services, violating both public expectations and the federal Canada Health Act governing federal transfers to the provinces for health care.

And that is precisely the (public) position of provincial medical associations, and (implicitly) of medical school representatives. Every service billed is provided, and every service provided is needed. (Otherwise it would not have been provided.) Therefore increasing billings simply reflect rising needs, and only that, and global caps impose on physicians the cost of meeting those needs. Increases in Q correspond to (significant) improvements in patient health.

But physicians’ representatives know perfectly well that their position is at best an overstatement. The whole raison d’être of the world-wide movement for “evidence-based medicine” is precisely the understanding, by clinicians, that much of medical practice is without a sound evidentiary base, and not a small proportion of it is contra-indicated by the evidence. For their part, provincial governments know that they cannot be sure, and certainly cannot prove, that all increases in billings are of no benefit to patients. And even if they could, it is virtually impossible for a government to win, in the public mind, an argument with physicians over medicine.

A serious effort to examine the evidence of health improvement from increased billings would be embarrassing for both sides. So physicians continue to make totally unsubstantiated claims of public benefit from increased billings, claims for which they are not held to account, and provincial governments continue to base policy on the opposite assumption, lack of health productivity, which they do not dare to articulate in public.

Representatives of medical schools, however, are in a somewhat more complex position. They are advancing arguments for increased physician capacity – increases in certain components of the N vector. That requires not only that more services are needed to maintain or improve health, but that more physicians are needed to provide them. Yet if output per physician (Q/N) is really rising, and this increased output is really contributing to patient health, as is asserted by physician representatives objecting to global caps, then perhaps the allegedly increased needs can be met without increasing the physician stock? Hence they must – and do – argue that for various reasons – feminization, changing attitudes toward work – physicians in future will put in fewer hours and less effort (Z/N down). They pass in silence over questions of increased productivity per hour (Q/Z up), so as to reach the desired conclusion of falling Q/N. But then, why are provincial governments trying to impose global caps on billings?13

The empirical evidence on productivity from the four countries at the Trilateral Conference is distinctly mixed, and everywhere incomplete. Some clear patterns, however, do emerge. In general, and despite all the speculation on changing physician attitudes toward work, there is little data on actual inputs of time and effort. Most of the measures reported are of output, variously measured, per practitioner – Q/N or bricks per Israelite. There are self-reported hours of work for Canada (QZREF) from 1982 to 1999; these show a pronounced U-shaped pattern. Hours (per week) spent in direct patient care fell during the 1980s, when physician supply per capita was still rising, and have since leveled off. Increases in the 1990s have been in activities other than direct patient care, possibly reflecting changes in definition rather than in activity levels themselves.

13 A possible reconciliation is that Q/N is rising in the present, but will fall in the future, and capacity must be expanded now to meet future needs. But these arguments have been made for over twenty years, and Q/N has continued to rise.
On the other hand, measures of output per physician show some quite interesting and contrasting trends. The Canadian data show a substantial increase in billing activity, at constant fee levels, for both medical and surgical specialists, but not for general practitioners.\textsuperscript{14} This is consistent with other Canadian sources that show fee-adjusted billings per physician to have risen quite rapidly over time for specialists, but to have changed much less (though with some up-trend) for GPs.

The U.K. data, by contrast, show falling output per physician, though again with more action among the specialists. Output per consultant, variously measured, seems to be trending downward over time, and procedural rates per senior surgeon in the NHS appear to be on average pitifully low, as well as highly variable from one surgeon to another. It may be argued that surgeons do other worthwhile things besides actual cutting, but it is also true that a surgeon with too low a procedural workload is unlikely to maintain adequate standards of quality. (For the individual surgeon, if $Q/N$ is too low, $H/Q$ falls and may in extreme cases become negative.) The rates reported for some surgeons surveyed seem to be in these dangerous ranges.

Moreover, even if the surgeons with low operating rates are doing other useful and important things, the degree of public concern over waiting lists and times suggests that they may have their priorities wrong. What is wanted, is surgery.

An obvious response to this situation -- highly variable and perhaps dangerously low workloads, and long surgical wait times, would be to take immediate steps to increase surgical productivity to levels more consistent with international practice and patient safety. The actual response has been to bemoan the shortage of physicians, and increase medical school capacity. On the surface (to the efficiency expert) this seems monumentally stupid. Behind it, however, are Pharaoh and the Israelites, with other political concerns -- in this case the lurking presence of private practice.

This is the political issue that neither physicians nor administrators -- nor politicians -- want to talk about. It is much more expedient to blame all the system’s problems on the eternal whipping-boy, the doctor shortage. (And to pass over in silence the data indicating that total medical and dental staff in UK hospitals, relative to the population, have nearly quadrupled since the inception of the NHS, rising 35% in the last decade alone.)

The productivity of British specialists may not be nearly as low as appears from the NHS data, because these do not include private activity. How much of the Z, the inputs of time, effort, and skill per practitioner, are being diverted to the private sector, where they may be highly productive -- and much better paid? Economically rational behaviour would consist of skimping on the supply of both hours and effort to the NHS, where one is paid on salary, and working much harder in private, fee for service practice. (Consultants have of course contractual obligations to the NHS, but these do not appear to be enforced with any rigour.)

\textsuperscript{14} The data are presented in a way that may not make this immediately transparent. Individual physicians are each classified according to whether their billing activity was above, within, or below “benchmark” levels for FTE status. From 1984 to 1996, the number of both medical and surgical specialists above the FTE benchmark range has risen faster than the numbers within or below that range, implying an increase in average billings (at constant fees) per physician. For surgical specialists the numbers within or below benchmark actually declined. For general practitioners, however, there is no obvious pattern, consistent with no significant change in productivity.
There are no incentives for consultant productivity in the NHS. Indeed there are counter-
incentives, since long waiting times push those who can pay into the private sector, where they can be charged fees whose level has attracted the attention of the UK Competition Commission. (Their finding of monopolistic price-fixing has not, however, been followed up.)

Data from the private sector are not, of course available to confirm or refute such speculations. It is worth noting, however, that proposals to buy consultants out of their “part-time” private practice with NHS salary increases of 30,000 pounds p.a. are thought to be far too low to attract those providing private elective surgery and diagnostic services. And the NHS productivity data are what they are. In Canada or the United States, by contrast, where surgeons are paid by fees for service, surgical productivity – or at least activity -- is high and rising.

The general practitioner story in the U.K. seems a bit different, though as Maynard and Bloor point out, general practice is a “black box” in which GPs do whatever it is that GPs do. The supply of GPs per capita has been rising steadily over time, and is now about 40% higher than when the service was established. The volume of consultations per GP on the other hand shows no clear trend, at least from 1975 to 1996. (It does, however show some quite pronounced year-to-year wobbles, that might raise questions of data quality.)

The impression of constant productivity per GP does, however, neglect the substantial increase in support staff in general practice. These numbers have more than doubled, from 1985 to 1997. Much of the increase has been administrative, but this should have freed up more practitioner time for patient care, and in any case the numbers of practice nurses have gone up nearly five times in this period. If output per practitioner has remained the same, this increase in non-practitioner inputs represents a substantial decrease in overall productivity – Z up and Q constant.

Consultations are not the only form of general practitioner output. There has been a deliberate NHS policy to encourage, through fees for specific services that depend upon the extent of population coverage, the performance of certain procedures or services such as immunizations and screening. These are presumed to be of particular importance for health maintenance, though Maynard and Bloor point out that the evidence of effectiveness is in some cases far from clear. Such evidence is of particular importance to justify interventions in healthy populations.

Indeed, they argue more generally that despite the presumed focus of the NHS on the health of the British population, there is virtually no attention given to the linkage from output to outcomes, from Q to H. But this is a much more widespread problem. Data from Australia confirm a pattern found in a number of countries – indeed everywhere that it has been looked for. The regional variations in (age-standardized) use rates for most medical procedures are wildly variable. Hospital separation rates for each of 14 common surgical procedures vary across the eight Australian states and territories, to an extent that is difficult to attribute to differences in underlying population need. Some of these procedures may be performed in facilities outside hospitals, skewing the comparisons. There is also an obvious pattern of markedly lower rates in the Northern Territory, which raises questions about access. But CABG rates are a third higher in New South Wales than in Western Australia, and Caesarian sections are nearly 40% higher in Queensland than in the Capital Territory. As in other jurisdictions, including the other three countries at this conference, such a pattern raises questions about the linkage between use and need, and feeds into the concern for “evidence-based medicine.”
Such variations data cannot, however, by themselves answer those questions. The relation between outputs and outcomes is no better addressed in Australia than anywhere else. Measures of output productivity, however, suggest that Australia may offer an interesting blend of Canadian and U.K. experience. The Medicare system reimburses physicians on a fee for service basis, as in Canada. Procedural output, or at least Medicare billings, per FTE practitioner shows a substantial increase from 1984-85 to 1997-98, and as in Canada, these increases are on average much greater for specialists than for GPs.

On the other hand, in Australia a significant number of specialists work on salary in hospitals, as in the U.K., and the hospital productivity data are much more ambiguous. Here, particularly for comparative purposes, it is critically important to be clear about the measures of inputs and outputs.

There is a widespread international trend of declining use of inpatient care, with falling lengths of stay and increased use of surgical and diagnostic facilities (whether or not attached to a hospital) in which the patient is not kept overnight. Since patient outcomes do not appear to be worse, and may in fact be better, with this new style of care, the change represents a clear improvement in productivity (Z/Q down). In both Canada and the U.S., inpatient days (per capita, age-adjusted) have been cut in half in the last twenty years. Australia has shared in this trend, as also has the U.K.

It is important to note, however, that these similar outcomes appear to have been reached through quite different processes. In both the U.S and Australia, governments adopted a payment system based on Diagnosis-Related Groups (DRGs) such that hospital reimbursements are based on the numbers of different types of patients they treat – a case-based “fee per item of service” system. In the U.K., “budget-holding” (now commissioning) general practitioners were made partially financially responsible for the costs of hospital care for their patients. The Canadian provinces simply tightened the global budgets of their hospitals, while fee-for-service physicians – both generalists and particularly specialists – maintained pressure to admit patients. All approaches succeeded – raising (but not answering) the question of some more fundamental and universal trend at work.

But what of physician productivity within the Australian hospital system? There is no necessary connection between shortening lengths of stay, and shifts to day care surgery, and increases in output per (salaried) surgeon or medical specialist. Since N, Z and Q are all vectors, one could have some ratios of elements of Z/Q falling – nursing hours or other hospital staff per treated case – falling sharply while others – physician time per case – were stable or rising.

As it happens, the input of FTE salaried medical staff per hospital separation has been rising over time, reminiscent of the U.K. experience, and is also highly variable across the states and territories. In the two years from 1995/6 to 1997/8, hospital separations per salaried FTE physician fell by 10.8%. This is not conclusive, as hospitals are also served by non-salaried Voluntary Medical Officers (VMOs). But if the increase in salaried FTEs is partially offset by a reduction in input of

---

15 This could change, however, in response to a recent court decision affirming patients’ rights to be informed of the skill level of different surgeons. The development of a set of risk-adjusted clinical performance standards for surgeons, similar to data generated a decade ago in New York State, would lead naturally into assessments of the appropriateness, as well as the competence, of surgical performance. But the process is likely to be very slow.

16 And as in Canada, the differences in productivity change across specialties are generating conflicts over the structure of fee schedules.
VMO time or effort, then presumably this time or effort is going elsewhere. VMOs may be putting more time into Medicare fee for service work outside hospitals, which would imply that reported productivity increase in that sector are over estimated. Or they may as suspected in the U.K. be doing more private work. The potential parallels with the U.K. experience seem to merit further exploration, and might warn against too ready an assumption of increased physician productivity in hospitals.

But there is one respect in which Australian experience quite closely parallels that of Canada, and differentiates thee two countries from the U.S. and the U.K. In both the U.S. and the U.K., there has been a significant and continuing increase in the numbers of “physician extenders” – in the U.S. both physician associates (PAs) and nurse practitioners (NPs), and in the U.K. practice nurses and related staff. Such personnel may work with or substitute for the physician; either way they permit an increase in medical services per active physician (or physician-hour). Changing the composition of the vector \( Z \), increases output per input of physician time. But the greatest productivity increases are achieved when intermediate-level personnel actually substitute for the physician. In both Canada and Australia these substitute personnel have been effectively shut out of the market completely. Whatever their priorities and intentions, these country pairs have come to very different places on one of the most important dimensions of physician productivity.

These very different patterns of personnel use reflect the differing economic incentives in these four systems, and demonstrate the very ambiguous role of productivity change. In both Canada and Australia physicians’ services are reimbursed through fees per item of service, but the reimbursement comes wholly (Canada) or largely (Australia) from a single public payer. In the U.S. fee-for-service has been the dominant payment mode (though it is declining) but payment is from a complex mix of multiple public and private sources. In the U.K. physician payment is predominantly from a single public payer, but is not fee-for-service. The combination of fee-for-service reimbursement with public single-payer funding seems to be particularly discouraging to productivity increase through the deployment of physician substitutes.

The issue turns on whether potentially cost-reducing substitutes become in actuality cost-increasing “add-ons”, and if so, at whose expense. In the hypothetical competitive markets of the economics textbooks, when a less costly input becomes available the inputs that it substitutes for either cut their prices, or leave the market, and the unit price of output falls. Medical “markets” are different.

The deployment of a significant number of physician substitutes, in any system, would represent a whole new group of income claimants – service providers who expect to be paid. Unless they are to work for free – or not at all – this implies either that physicians collectively must be paid less, or total payments for medical services must rise. But in a single-payer system, the latter creates problems for government budgets, while the former is likely to be unacceptable to well-organized medical associations bargaining on their members’ behalf.

---

17 Whether this is “efficient” depends upon the relative cost of such substitutes, compared with the extent to which they increase output per physician.

18 Domestic arguments over the feasibility of such substitution can be dismissed as simply blocking tactics, not serious questions. The research was all done long ago, and was quite conclusively positive, and such personnel are “in the field” and working in other jurisdictions.
In principle, if the numbers of physicians could be reduced simultaneously with the increase in substitutes, average physician incomes could be maintained or even increased while total public outlays stayed constant or fell, on the presumption that the physician extenders really are less costly substitutes. In practice, it will be difficult to find volunteers willing to leave practice prematurely – to give up their incomes – so as to make space for their substitutes. The competitive market of theory just pushes people out; in medicine and particularly fee for service medicine, that is difficult or impossible to do. However attractive in principle, in practice the physician-extenders have no (significant) friends. Paying governments reject them if they add to overall costs, and physicians reject them unless they do.

So how is the U.S. different? Well, historically it was not so different, until the (apparent) advent of genuinely competitive managed care. Organizations that are under competitive pressure to hold down the total costs of the care they provide, in order to attract enrolled members will and do look for less costly substitutes, and their numbers are growing.

But what happens next? Do the physicians let go by managed care organizations become unemployed, or go into real estate? Do they accept lower incomes, to stay competitive? Do they find ways, in the remaining fee for service segment of the market, to generate additional activities (selling herbal remedies, or Amway products, for example? Or do they engage in a combination of public relations campaigns and political action to discourage prevent managed care companies from making these cost-reducing substitutions?

At present, anecdotes indicate that all of these responses are occurring, but it seems too soon to judge either their relative magnitude or their effect. What seems most likely, however, given the fragmented nature of the American health care funding system, is that total national expenditures on health care will rise to absorb all of the additional personnel. The mix, however, may be different in different sectors.

The question is particularly salient, however, with respect to the increasing supply of female physicians. So far, females have displayed a marked preference for what in the U.S. are known as “primary care specialties” and for the aspects of practice – an emphasis on interpersonal relations rather than the technical and procedural. They tend also to be less “entrepreneurial”, more likely to in salaried practice. These are the areas most open to substitution of PAs and NPs. There thus seems to be, in the U.S., a looming conflict over “market share” between female physicians and physician extenders – precisely the conflict that physicians in Canada and Australia have foreseen and moved to prevent.

To sum up, we return to the Figure showing the chain from capacity (to produce medical services) to outcomes (from their use). What may we say about each link in the chain?

Starting at the bottom, there seems to be consensus across all countries that little, certainly not nearly enough, is known about the relationship between outputs and outcomes. The former are measured quite well in some countries (e.g. Canada) and rather badly in others (e.g. the U.K.), but the latter are measured very badly everywhere. It is thus not surprising that there is no good evidence, at the system level, of the “productivity” of health care, either in itself or relative to other social and economic factors, in producing health.
Great strides have been made, both in methodology and application, in measuring the impact on health of particular care interventions, but typically under experimental conditions. To move from this to determining the effect of increasing or decreasing care in aggregate at the population level, however, is a much more difficult task. We might think about this question by defining four categories of people:

1. Those who might expect to benefit from a particular set of treatments, and are receiving them;
2. Those who might expect to benefit from a particular set of treatments, and are not receiving them;
3. Those who are receiving a particular set of treatments from which, on the basis of current scientific knowledge, they cannot be expected to benefit;
4. Those who neither need, nor are receiving, health care services.

These correspond to effective care, unmet needs, ineffective or harmful care; and incorrigible health, respectively.

A person may, however, fall into more than one of the first three categories. Someone receiving effective but non-optimal therapy could be both in categories 1 and 2, for example, while someone receiving ineffective or harmful care would be in category 3, but might also be in category 2 depending upon whether or not an alternative, effective therapy was available. Over-treatment would be covered by categories 1 and 3. The terms “appropriate/inappropriate” are useful precisely because of their imprecision; they cover a number of different cases.

The “productivity” of health care with respect to health is increased, if people are moved from category 2 to category 1 – meeting needs – and out of category 3 – eliminating unnecessary care. Much of the conflict over the appropriate scale of a country’s health care system can be reduced to disagreements over the relative sizes of categories 2 and 3, and particularly over whether increasing the output of medical services (by training more physicians, for example) would tend to reduce 2, or increase 3.

The question of the impact of an increase (or decrease) in service output cannot, however, be divorced from one’s interpretation of the current state. And that is why it is so threatening. If, in a given system, category 3 is large, then the obvious way to increase productivity is to eliminate unnecessary and ineffective services, and scrutinize the unproven. Alternatively, if 2 is large and 3 small, then the health care system is providing appropriate care, but there are many unmet needs. The former scenario would be quite embarrassing for physicians and other service providers – they are doing, and being paid for, a good deal of care that they should have known (on the basis of currently available evidence) to be useless or harmful, or at best unproven. The second scenario, however, is similarly embarrassing for governments, who are ultimately responsible (even in the U.S.) for the functioning of the health care system. Why are these needs going unmet?

There is a good deal of partial evidence to suggest that category 3 may be quite large, in all the countries represented. The high rates of variation in procedural performance certainly point in that direction, though they can also be read as evidence of category 2 – as in Australia’s Northern Territory. But no one has tried to assemble the results emerging from studies of the effectiveness of particular interventions up to a system-wide evaluation of the productivity of current practice.
patterns. Reliance on emphatic and unsubstantiated claims buttressed by lurid anecdote is much safer, and effective for most purposes.

We do have a clear example of a major advance in health productivity in all four countries, occurring in the hospital sector, not in medical services. Hospital utilization, measured in (age-adjusted) patient-days per capita, has been trending down for several decades, but the fall has accelerated in the 1990s. Lengths of stay are down, many more patients are being cared for without an inpatient admission, and there has been no deterioration, in fact more likely an improvement, in patient outcomes. In the schema of Figure 1 above, and depending on whether we regard inpatient days as an output, or as an input to the production of treated cases, we may call this a reduction in Q without reduction in H, or a reduction in Z without reduction in Q. Either way, it is an unambiguous and substantial improvement in productivity.

The history of this process, however, illustrates three points of significance for the discussion of medical productivity.

First, it did not come about through any advance in technology, or other form of “know-how.” Every student of hospital utilization, since at least the 1960s, has pointed to excessive lengths of stay and over-use of hospital beds. The experience with day care surgery, in Canada, the U.K. and the U.S., dates from the late 1960s and early 1970s, and is readily available in the published literature. But sheer organizational and professional inertia held this process back for a decade or more (depending on the country) with the result that hospital costs throughout the 1970s and 1980s were far higher than they needed to be.

Contrary to the opportunistic claims by pharmaceutical manufacturers, it was tighter financial pressures, not new (and expensive) drugs, that brought about the change. Productivity improvement depended not upon the discovery of new knowledge but upon the application of old knowledge – on behavioural change forced by external circumstances.

Second, in Canada in particular it is quite striking that there has been a slow decline in hospital use (measured in patient-days) since the 1960s, with no corresponding decline in hospital budgets. Only in the 1990s do hospital budgets fall (per capita), along with the accelerating decline in patient-day use. These declines have been accompanied by a dramatic escalation in the public rhetoric of system collapse. It is only when budgets shrink, that people lose jobs and incomes. So long as budgets are maintained, a reduction in one form of (paid) activity can be made up by an increase in another. A similar process seems to have occurred in the U.S.; the decline in inpatient use in the 1980s led to a shift of costs from inpatient to ambulatory settings, but decline (or rather reduced rate of escalation) had to wait for the managerial re-organizations and changed incentives of the 1990s.

And finally, despite overall reductions in patient-day use of from one half to two thirds, we still find significant variations across regions in the rates of performance of particular procedures. One might have expected these rates to converge toward some level of objective “need”, and they may yet. But

---

19 The shift in activity patterns may of course represent an improvement in productivity, if resources are being shifted from ineffective to effective forms of care. But the point is that that is not the point. The shift takes place to keep the resources occupied – and paid – it is not generally accompanied by, and does not depend on, evidence of effectiveness of the new activities.
so far, in the U.S. environment where hospital use rates are lowest, the relative differences across regions seem to be preserved.

The experience of productivity increase in hospitals thus illustrates the importance of the conflicting objectives involved. The constraints on productivity were behavioural, not technical – like Pharaoh’s refusal to provide straw – and the inhibiting behaviour had roots in good solid economic interest. Breaking down those constraints required external pressures; these could be applied in several different ways. “Managed care” in the U.S., budget-holding in the U.K., freezing of hospital global budgets, and merging or closing hospitals in Canada, each forced change by reducing the flow of resources to hospitals.

This is not to say that technological progress in therapy plays no role in improving productivity. It would not be difficult to compile a long list of technical improvements that have reduced the invasiveness and side effects of particular treatments, and reduced the need for hospital care. But the historical record suggests that these advances play an enabling rather than a precipitating role in productivity improvement. Improvements in individual procedures do not add up to system-wide changes, and the opportunities they create for global productivity improvement, are not taken up in the absence of pressures from the external environment.

The importance of the pressures or incentives generated by the external environment is again illustrated when we return to physician productivity, and consider the relation between the outputs Q and the inputs Z – as often proxied by N.

The evidence in these papers seems consistent with the predictions of the most elementary ideas about economic incentives. Output per physician, as represented in billed services and procedures, tends to rise over time in fee for service environments but not in salaried or capitated ones. There is of course more to the story. The increases in fee for service systems may not be (wholly?) real, arising instead from “creative billing”. Productivity in salaried systems may be underestimated if consultants are shifting their time and effort to the private sector – again a straight-forward prediction of the elementary economics of self-interest. Technology matters; productivity seems to rise more rapidly in the procedural specialties than in the “cerebral” specialties and general practice. But the simple story, that productivity responds to payment mode, seems to hold up internationally.

On the other hand increases in productivity per physician that are not captured as income by physicians, are much less likely to occur. It has long been known that output per general practitioner in particular can be greatly expanded through the deployment of various categories of “physician-extenders”, but the gains from this increase would not necessarily accrue to general practitioners themselves. Where such substitutes are more likely to be a threat to physicians’ market share and incomes (Canada and Australia), they are virtually non-existent. Where GPs can capture some or all of the gain, they are. In the U.S. the regulatory and payment environment is so complex that it may not be possible for “organized” medicine to protect its market share; on the other hand it may also not be possible for payers to avoid a general expansion in income claims and expenditures.

Going to the top of the chain, an important reason for being interested in physician productivity in the first place is presumably as an input to the planning of physician supply and training capacity. If we believe that there is some objective meaning to the idea of “need” for physicians’ services, however difficult it may be to define in practice, then we will have to decide, either deliberately or by
default, how many physicians are required to meet those “needs”. And that decision embodies explicit or implicit assumptions about productivity per physician.

But the concern for the rate of inputs per physician – Z/N or working hours per trained practitioner -- while logical in principle seems less relevant in practice. Claims have been made for years that changing life-style aspirations and the feminization of medical practice were leading, now or soon, to a decline in Z/N and a need for more physicians.

But “life-style aspirations” have more than one dimension. Physicians might, quite understandably, wish to work less, but are they also willing to earn less? Physicians’ average incomes are at the top of the list of occupations and professions, and high incomes tend to go along with high work-loads. Judged in terms of Adam Smith’s concept of “net advantages”, medicine remains (on average) a most, perhaps the most, attractive occupation, as is reflected in the intense competition for entry.20

A reduction in physicians’ income aspirations has not, to date, been widely apparent; until that happens we are unlikely to seen changes in the Z/N reflected in reductions in Q/N. What we may see in fee for service systems, is changes in patterns of practice and billing in primary care that permit higher levels of income for lower levels of time and effort, and effective care. A number of physicians seem able to describe in very specific terms how this can be done; this raises fundamental questions about the appropriateness of fee for service reimbursement in primary care. But such questions then take us back to the relation between output and outcome, where again we are confronted with firm assertion and little information.

At the level of the relation between physician capacity and input supply, there is an obvious asymmetry in the public debate that is highlighted by the papers at the Trilateral conference. Great attention is given to projecting hypothesized declines in Z/N; much less to assessing the sources of change in N – in particular the effects of international migration. In some countries – the U.S. for example – this component of capacity growth seems to be beyond policy control, but in others – Canada, for example – changes in immigration policy have had very large and rapid effects on the supply of immigrant physicians. In any case one might want to predict these shifts with some care. In all countries represented at the Trilateral conference there is tremendous pressure from foreign physicians attempting to enter practice.

Medical schools are in the capital equipment “business”; they produce and sell the form of human capital known as a medical education. In marketing this product they understandably emphasize the rate of depreciation of the existing capital stock, and dismiss or denigrate the possibility of a country’s acquiring that stock, either deliberately or inadvertently, from other sources.21 Accordingly their representatives tend to focus on and perhaps overestimate out-migration and underestimate or ignore in-migration. Representatives of Canadian schools, for example, give great publicity to the very substantial out-migration of their graduates to the U.S.

20 Those who claim that a career in medicine is not as attractive as it was may well be right, but the same seems to be true across the board.
21 The analogy is not perfect; manufacturers of physical capital do not generally try to increase sales of their products by predicting a decline in their productivity! [We must train more doctors, because each one we train is going to work less in future.]
Yet Barer and Webber show that, despite substantial efforts to keep them out, in-migrants to Canada more than balance this out-flow. We are net gainers of physicians from international migration. Furthermore, changes in immigration policy now being considered, for reasons totally independent of physician supply or needs, could have the effect of dramatically increasing this in-flow. Similar shifts in the external policy environment, arising from the commercial integration of the European Community, could have significant effects in the U.K.

Under these circumstances, countries should be most reluctant to make very long-term commitments to increases in physician supply. If policy changes in the external environment, that health care planners can neither predict nor control, may have major effects on capacity, then one should be very suspicious of long-range projections of “needs” for domestic training capacity. All predictions will be falsified; it is best to stay light on the feet and be very alert to changing circumstances. The worst policy is likely to be one that locks in for decades a particular structure and growth path for the physician work-force. We did that in Canada, thirty years ago, committing ourselves more or less by accident to a doubling of (per capita) physician supply, and are only now regaining our freedom of action. We may yet make the same mistake again.

This leads back to physician productivity. If our concern is for the availability of medical services, then the international evidence suggests that output per physician is highly variable both within and across countries, and over time. Productivity is very responsive to the incentives and constraints of a particular organizational environment. Technical parameters presumably set some broad limits to these responses, but if so, those limits appear to be very broad indeed. The comparison of actual experiences highlights the importance of the behavioural factors, not the technical ones.

Or to put it bluntly, we already know several ways to make physicians much more productive than they are now, “we” (i.e. the net resultant of all the conflicting interests involved) just choose not to do so. The same was true of hospitals, twenty years ago. The apparently “objective” and technical questions of measuring and promoting physician productivity, turns out instead to be a conceptual and statistical battleground in a broader conflict over income shares, global expenditure levels, and control of the style and content of medical practice itself. It would be so much simpler if we were making bricks…

22 “Marry in haste, repent at leisure.”