

A New Paradigm for Health Economics? We Already Have Three!

ROBERT G. EVANS *

Revised, February 2012

Draft presented at the 8th World Congress on
Health Economics of the International Health Economics Association (iHEA)
Toronto, July 2011

The inter-sectorial financial flows that characterize all modern health care systems can be represented compactly in the accompanying Figure I. Adapted from the standard National Income Accounting framework, the figure shows the principal “pipes” through which financing flows from the households comprising a nation’s population, to the intermediate agencies (governments, social and private insurers) that assemble and re-distribute collective funds to the various types of provider organizations (public and private clinics and hospitals, and for-profit commercial firms) that in turn produce the multiplicity of different health care goods and services. These firms then distribute their funds, directly or indirectly, back to various persons as payment for their “factor inputs” – labour and management skills and the services of various forms of capital. A fundamental feature of this framework is that ALL the financial flows originate with households and return to households – though not of course the same ones. All revenues raised become expenditures, which in turn all become someone’s income.¹

In the heart of the framework are the “real” flows, the quantity and mix of different health care goods and services that are produced by firms, using the factor inputs supplied to them by households, and which flow back to households – though again not in general the same ones.

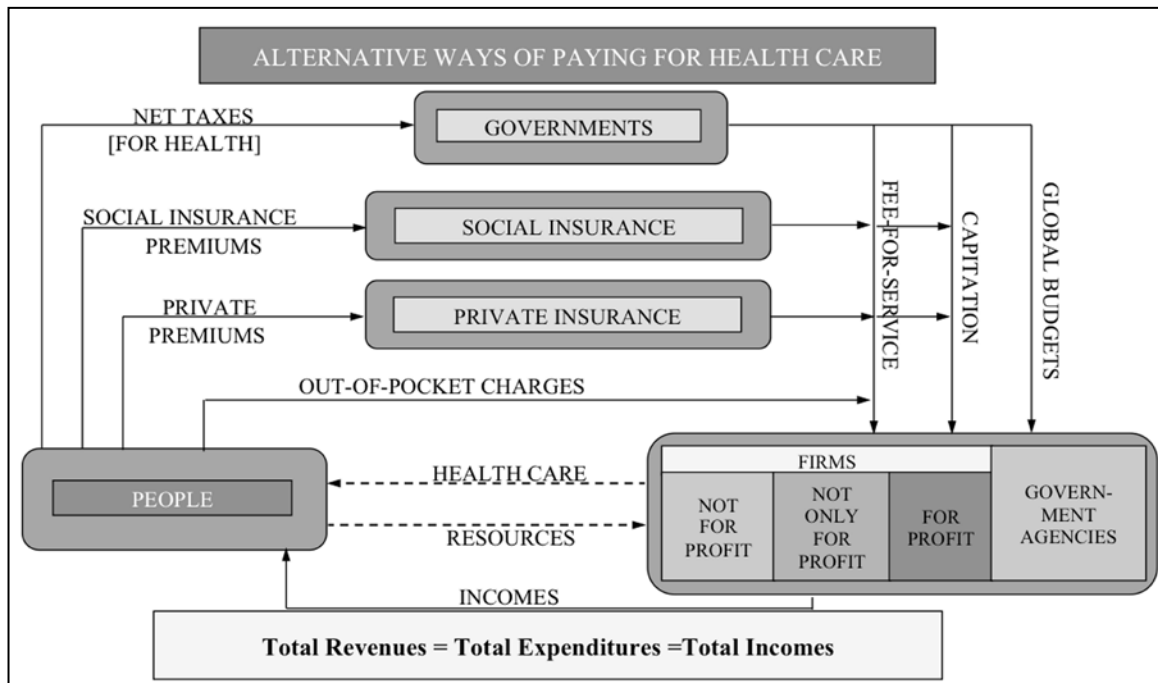
But Figure I is a set of accounting relationships, a gross anatomical description that provides no “physiology” to explain how the various components interact, or how those interactions might change in response to anatomical changes. What difference does it make, in terms of patterns of service delivery and cost, of distribution of burdens and benefits among the population, or of population health status, if the mixes of financing and funding flows are re-arranged? These questions, sometimes overt, often covert, are everywhere at the heart of debates over health policy.

Attempts to develop such a “physiological” understanding, however, are impeded by the embedded conflicts of interest associated with any policy choice. Research and analysis, no matter how objectively motivated and scrupulously conducted, will if acted upon always have distributional implications, often readily apparent. Representatives of threatened interests will challenge the analysis and propose alternative interpretations of “how things *really* work”. These cross-cutting motivations radically compound the

¹ Because this Figure is adapted (much simplified) from the general National Income Accounting framework, we can draw on that much broader framework to reconcile “missing bits” such as the absence of a foreign sector, debt flows, or capital accumulation. They can all be fitted in!

ambiguities inherent in trying to understand the behaviour of any complex system. Accordingly there is no single settled body of physiological understanding of health care systems to correspond to the more readily observed anatomy.

Figure I



One can, however, identify three quite distinct perspectives or “frameworks of understanding” through which people – clinicians, patients, researchers, government officials, politicians, and the general public – interpret the behaviour of health care systems. These we may label:

- The “Naïve Clinical”,
- The “Mainstream Economic” and
- The “Eclectic Structuralist”.

Casual empiricism suggests that the Naïve Clinical is by far the dominant perspective among both clinicians and the general public, and typically exerts its influence over health policy through the sensitivity of politicians to that professional and public opinion. The Mainstream Economic perspective, by contrast, is held almost exclusively by professional economists or those with conventional economic training. It is predominant in this group (though not among those who specialize in health economics) and derives its influence from the strategic positions they occupy in government economic ministries, in the corporate world, and in the business press. Finally the Eclectic Structuralist perspective tends to be held by health services researchers (including most health economists), officials in government ministries responsible for health or health care, and administrators/managers of health care institutions (including former clinicians).

These three perspectives – paradigms – each postulate a different pattern of causal connections or perhaps better patterns of influence among the accounting relationships in Figure I, and their relationship to the health of individuals and populations. They are each composed of three distinct elements:

- (1) A *normative* view of how levels and patterns of health care utilization *should* be determined,
- (2) A *positive* view of how, under appropriate conditions, they *are* determined, and
- (3) A corresponding set of stories that focus research, analysis, and policy on a particular sector of the circular flow above – the sector that “really matters”.

Each perspective also has its characteristic silences, gaps where no plausible story is told about particular linkages in Figure 1.

The distinction between the positive and the normative components of these different perspectives is of central importance for the policy recommendations that emerge from each. As all first-year students of economics are (supposed to be) carefully taught, a positive statement is a statement – true or false – about facts. “Water (pure) boils (at sea level) at a temperature of 100° Celsius,” or “The moon is made of green cheese.” In effect these are predictions about the outcome of particular observations, to be made under more or less completely specified conditions.

Normative statements, by contrast, contain the words “ought” or “should”, either explicitly or in some rhetorical equivalent. They assert that one state of the world is to be preferred to another – is in some sense “better” – and typically include an implication that someone – not necessarily clearly identified – should act to ensure that the better state prevails. The almost universal claim by health care providers that “health care (in our sector or system) is underfunded”, for example, is equivalent to “More money should be spent on health care” – a claim that the world would be a better place if expenditures on health were larger than at present. The total amount of money flowing around the circuit in Figure I should be larger. (Note that those making this claim are often unclear as to whether they advocate increases in some components of health services, or in the amount paid for them – more care, or higher pay rates.)

One may or may not share the preferences expressed in this statement. But the key point is that it *is* a statement of preferences, a statement about the relative values that the speaker attaches to two different states of the world (one with present spending levels, and one with some higher level). It is not a statement about the world, but a statement about the speaker’s attitude toward the world, about her values.² The speaker will typically be trying to create and/or mobilize a broader constituency of people who share those values, in order to increase the chances of bringing about the desired change – more money spent on health care.

A common way of building support for change, and neutralizing opposition, is to present the normative statement as if it were a positive one – *e.g.* “There is a shortage of doctors, and that’s just a fact.” The number of doctors available to serve a particular population is, subject to questions of definition and measurement, a potentially confirmable or falsifiable fact. So is a claim about the “needs” of a population, in terms of

² Even the most distinguished scholars can occasionally be trapped by the language. In the course of a very clear and powerful exposition on this subject (which deserves reading by every health economist and most of the rest of the profession) Fuchs (1996) carries out an “unscientific” but very suggestive survey of colleagues’ opinions as to the validity of a set of positive and normative propositions about health and health care. But one of his “positive” propositions is actually normative. (Its identification is left as an exercise for the reader.)

the expected impact of different treatment levels and patterns on health. (The word “need” itself, however, is ambivalent insofar as it asserts an obligation on some other(s) to respond to that need.) But the language of “shortage” is clearly normative, implying that “something should be done” – perhaps open more medical school places – because if there were more doctors, the world would be a better place.

The fundamental logical distinction between normative and positive propositions was spelled out by David Hume in the 18th century: “Hume’s Law” is that one cannot (logically) derive “ought” from “is”. No assembly of valid facts, or of confirmed positive propositions, can serve as a basis for normative claims – or policy recommendations. There has to be brought to that assembly of facts a corresponding set of values and preferences with which to rank alternative states of the world. It is always a logical fallacy – sometimes inadvertent, sometimes deliberate – to claim to base recommendations on “science” – or for that matter “common sense” – alone. But the point deserves emphasis, because that fallacy is so frequently committed, even by professional economists.

The normative component of the clinical perspective is simply that people should get the care they need, as judged by a qualified clinical practitioner, regardless of the cost. The criterion for whether resources should be allocated to produce a particular form of health care, and whether that care should be offered to/accepted by a particular patient is: “On balance, and allowing for uncertainties, is this intervention likely to do more good than harm to the patient’s health?” If the answer is believed to be “yes”, then the patient needs the care and that need should be met.³ This ethical norm is so deeply embedded in professional practice that it is probably no longer recognized as a particular ethical position, among other possibilities. It is simply what you do.

Yet the mainstream economic perspective rests on an alternative normative basis, “consumer sovereignty”, that could hardly be more different.⁴ People *should* get whatever care they are willing and able to pay for, at prices reflecting the resource cost, the real opportunity cost, of producing that care. The impact of that care on their health status is irrelevant, a position symmetric with the disregard of cost in the clinical norm. “Consumers” (not patients!) presumably take account of health effects, among other things, in choosing how to spend their incomes, but if they want and are willing to pay for useless or even harmful care, then they should get it. Conversely, if they are “unwilling” – which includes unable – to pay for desperately needed, even life-saving, health care services, then they should not get those services.

It is “allocatively inefficient”, from this perspective, for a society to allocate resources to producing services that people do not value sufficiently to be willing to pay the full cost of producing them. The world will be a better place if those resources are used instead to produce services that someone *is* willing to pay for (Pauly, 1969; Feldstein, 1973). This sounds plausible, until one recalls that “value” in this context refers not just to intensity of preferences but also to ability to pay. “Them as has, gets, and them

³ There is a certain ambivalence on the question of disagreement between patient and professional-- what if the offered care is rejected? In principle (and in law) the wishes of the (competent, informed) patient are supposed to be determinative. In practice, however, the clinical perspective seems to include a belief that the patient who acts against medical advice is either not fully informed or not fully competent. Patient non-compliance with drug therapy in particular is typically viewed as either a failure of communication by the prescriber, or an inability of the patient to remember and follow instructions. Such deficits should be corrected.

⁴ “Consumer sovereignty”, like “need”, has both a normative and a positive sense. It can mean, as here, that people *should* get what they want, or that in fact they do.

as hasn't doesn't" is not just a blunt observation about reality but a moral principle – the way things *should* be in the best of all possible worlds. If there is a change in the income distribution, however, then the definition of the “right” pattern of resource allocation and commodity distribution shifts to respond to the new distribution of ability to pay.⁵

Thus both clinical and economic perspectives include well-defined (in principle) concepts of appropriate and inappropriate health care provision. But their concepts are very different. To the clinician, care provided to a particular patient is inappropriate if it is unlikely to do more good than harm for that patient's health. From the mainstream economic perspective, care is inappropriate if the patient would/could not pay for it out of his/her own resources (or voluntary donations by others).

Reinhardt's (1998) hypothetical example of the wealthy Changs and the impoverished Smiths is particularly instructive. His point is not simply that in the American context the Changs will (if they choose) receive for themselves and their children all the health care they want, including care that is of minimal (if any) benefit, while the Smith may have to forego care of considerably greater potential benefit, putting their health at much greater risk. This is true, but his key point is that much if not most of the formal analysis of health care in the economic literature implicitly adopts as a normative presumption, a moral principle, that this is the way health care *ought* to be allocated (see also Reinhardt, 1992).

Conflicts of moral principle cannot be resolved by rational argument; ultimately they are matters about which people can only fight (or vote).⁶ But the normative underpinnings of the mainstream economic perspective set up some extremely interesting cognitive dissonance among economists.

First of all, it is not clear that many economists, if any, actually accept, explicitly and whole-heartedly, this normative position in their non-professional lives. A surprising number even of American economists, replying to the informal survey described by Fuchs (1996), expressed the view that people should not be denied needed health care because of inability to pay, and that view may be much more general. Only in their academic work do they seem committed to a framework of analysis that rests on the alternative ethical principle. Arrow (1976), perhaps recognizing the rather squalid moral basis for his analysis, states: “In order to avoid distributional considerations, I shall assume that the economy consists of a single individual.” He neglected to remind readers that any resulting findings were relevant only on Mars.

But if this is so, the internal ethical conflict seems to have been resolved in a very peculiar way, through the belief that the normative position embodied in the economic perspective is somehow part of the discipline of economics itself, either derived from economic analysis or in some other way inextricably connected with “doing economics right.” As a professional economist, one has to adopt this moral position, otherwise one is

⁵ The normative position underlying the Mainstream Economic perspective is independent of how the distribution of ability to pay comes to be whatever it is. Most people consider theft as an immoral form of redistribution; some of the more extreme advocates of “free markets” seem to regard tax and transfer systems established by duly elected governments as equivalent to theft. On the other hand any redistribution occurring through a market transaction (for which no one has been convicted), including some of the amazing transfers in capital markets, presumably makes the world a better place. In any case the allocation of resources should certainly respond to that redistribution. But here we enter the realm of “Natural Law” theology.

⁶ Better information about the consequences of adopting different positions may eventually influence people's normative views, although unfortunately normative positions also have a strong influence on what is recognized as valid information. “I wouldn't have seen it if I hadn't believed it.”

not doing “real economics”. Yet as noted above, Hume’s Law points out the logical fallacy of imagining that one can derive “ought” from “is”. It is logically impossible for economic analysis to generate the normative presumption that people should get only what they are able and willing to pay for. That principle has to be annexed from outside, as a moral choice by particular individuals.

A group of such individuals – perhaps mainstream economists – may adopt as a convention that “real economics” must include a set of moral principles that they personally find congenial, or perhaps simply convenient for analytic purposes. They may then claim that “economics itself”, as they define it, imposes on its practitioners the acceptance of those moral principles. This appears to be the view of Pauly (1996), replying to Culyer and Evans (1996), in which he actually apostrophises “Economics” as some transcendental entity, demanding that its practitioners accept particular assumptions. But at this point one has crossed the border from an academic discipline to a religion. It is priests, not academic analysts, much less scientists, who are in the business of dictating the normative principles that one is required to hold, to belong to a particular community.⁷

There is not much interesting to say about the normative component of the “Eclectic Structural” perspective. In general its practitioners seem to share the clinical position that people should get the care they need. Certainly health services researchers spend a good deal of time and effort studying the impact of different forms of health care on health. But they might argue that this normative view is grounded in the (casual) empirical judgement that this is what the general public (including most economists in their private lives) want the health care system to do (e.g. van Doorslaer et al., 1993). Normative principles should be derived from peoples, not priests (though that itself is a normative judgement). The normative position underlying the clinical perspective, by contrast, appears to be more of a “categorical imperative” for clinicians, not ultimately derived from broad popular or even legal support. Such a moral position may be more priestly than scientific, but most clinicians probably find that dual role quite acceptable.⁸

The positive component of the economic perspective is imported directly from standard textbook economic theory of “widgets”. People will in fact get the care they want and are willing and able to pay for, at prices reflecting its real resource cost – “consumer sovereignty” in its positive sense – if they can/must purchase it freely, at their own expense, in perfectly competitive markets supplied by for-profit firms. Competition among such firms will ensure that production is technically efficient – no wasted resources – and carried out with the optimal – lowest cost – mix of resource inputs, while the requirement that “consumers” must pay for their own care out of pocket ensures that it goes to those who value it most – as indicated by willingness to pay.

The questions of who pays and who gets paid for care are in this context easily settled: the market will decide. Users of care will (subject to the one qualification below) pay the costs of their own care, while the mix and rates of earnings of factor inputs – the rates of pay of doctors, nurses, and other professionals, and the profits of provider firms, and all the political controversies that vex public systems of health care – will be

⁷ Participants in this convention may, if they are able to occupy strategic positions in university departments, academic journals, and research funding bodies, make it very difficult in some countries to “do economics” without accepting their preferred moral principles. The validity of the definition of true doctrine offered by St. Vincent of Lérins: “*Quod semper, quod ubique, quod ab omnibus creditum est*”, rests implicitly on the power to burn those who disagree.

⁸ Clinical norms, however, focus more on the individual patient. “People should get the care they need” is a guide for general health policy and financing. But *my patient* should get the care she needs, even if that implies denying care to others in much worse case.

determined in the decentralized bargaining of competitive private markets. And the answers that emerge will, by definition, be the "right" ones.⁹

The only difference from the standard textbook account of the production and consumption of "widgets" is recognition that individuals' "demand" for health care is subject to random and sometimes quite large fluctuations. One could treat this as fluctuations in "tastes" for a particular commodity, health care, but large and rapid changes in consumer "tastes" might undermine their normative plausibility – why should giving consumers what they want be an ethical principle if their wants are so unstable? Besides, to describe an individual's response to a heart attack or the onset of cancer as a sudden "change in tastes" sounds (and is) artificial and stupid.

Implicitly, then, the mainstream economic perspective has to adopt a framework in which health status is subject to random shifts that both lower individual well-being and raise the perceived health payoff from "consuming" health care. For a previously healthy individual in a serious car crash, for example, or diagnosed with cancer, "tastes" for care would jump discontinuously and very suddenly from near zero to some very large value. The rational consumer reorders her consumption patterns accordingly – (much) more health care and less of other things. But the very best that can be achieved from this increase in health care use – rarely achieved in the case of serious illness or injury – is to restore the health *status quo ante*, while the reduction in income available to purchase other commodities and the discomfort and distress associated with the care itself cause a drop, possibly very substantial, in the individual's overall well-being.

By incorporating the experience of illness and the purchase of health care into standard consumer theory in this way, however, the economic perspective introduces another and very important pair of implicit assumptions. In order for the rational consumer to carry out the reordering of consumption patterns required to maximize utility under the new circumstances, she must be in operational as well as strategic control of the "consumption" process – a highly dubious assumption, particularly in the real circumstances of the serious illnesses that account for most of health care use and costs. More fundamentally, however, she must know the relevant marginal utilities. She must not only know how much she values health – the standard assumption in consumer theory – but must also know, if not perfectly then at least better than anyone else, the impact in present circumstances of health care on health status.

For this assumption there is no warrant or precedent in the economic theory of the consumer. In the community at large, including both patients and providers, the standard and seemingly self-evident assumption is that professional experts usually have better, typically much better, information about this relationship. Information is asymmetric between provider and user of care; professions know patients' needs better than patients do. The Mainstream Economic perspective simply assumes away this asymmetry, typically implicitly, and provides no justification for doing so.¹⁰

⁹ If competition among firms or factor suppliers is impeded by regulations or collusive behaviour, then the obvious answer from the economic perspective is aggressive public policy to promote competition, in health care as everywhere else. *Inter alia*, remove all self-regulatory powers from the professions along with any other barriers to entry. Some of the more extreme market advocates seem to go farther, however, to suggest that even *this* form of public intervention in private markets is misguided. Perhaps "the market" somehow produces an optimal level of monopoly and collusion?

¹⁰ In some analytic frameworks, such as "health capital stock" models in which people are assumed to "invest" in accumulating "health capital" by *inter alia* consuming health care, the assumption is made explicitly that individuals have full, or at least privileged, information about the structure of the (sometimes rather peculiar-looking) health production function. Other frameworks conceptualize the consumer as a

The individual uncertainty created by the possibility of these random fluctuations in health and well-being is, however, assumed to be a “bad” in itself, and people are willing to pay to be relieved of this bad by purchasing “health” insurance that provides monetary compensation for loss of health. Where there is a demand, there will be a supply. Private firms will offer such compensation, in competitive insurance markets, at the “right” price. But there is no market in which one can buy or sell health, and accordingly no way of insuring it. Health insurance subsidizes health *care*, by enabling “consumers” to acquire and use it below cost – perhaps even free. For this, each individual will be charged a premium proportional to his or her risk status, hence individuals continue to pay for their own care in a probabilistic sense.¹¹

At this point the economic perspective immediately detects an allocative distortion – an inefficiency. If any commodity is available to consumers at a price below its opportunity cost, they will use “too much” of it, withdrawing resources from more highly valued uses. There is thus, from this perspective, a trade-off between the increased well-being from being able to reduce the “bad” of uncertainty by in effect selling it to the insurer (where it is pooled and disappears), and the loss of well-being from the “overuse” of care in the sense defined above – a definition of overuse in terms not of the effectiveness of care, but of individual willingness/ability to pay.

The existence of such a trade-off is one of the fundamental features – perhaps the fundamental feature – of health care financing as viewed from the mainstream economic perspective. Yet unless one accepts the normative principle, the value premises, underlying the mainstream analysis – which few if any do outside academic economics – the “trade-off” vanishes into thin air.

Nevertheless, this imaginary trade-off looms large in the academic literature, particularly in the United States. It has been particularly effective in focusing mainstream economic research on issues of interest to the private insurance industry. A great deal of research effort has been devoted to exploring its terms and suggesting ways of structuring insurance systems so as to minimize, on some metric, the combined costs of uncertainty and allocative distortion. Most of this academic research, however, seems to point (as did Arrow, *op cit.*) to the inescapable conclusion that optimizing this trade-off requires some combination of individual and third-party payment – precisely what most public systems of health insurance do not include.¹²

The mainstream economic perspective thus focuses attention on the upper left branch of Figure I, the mix of financing channels, and particularly on the level and form of self-payment in the total revenue mix. Since decisions to consume health care are assumed to be made, like any other consumption decision, by individuals responding to the prices they face, the financing mix determines through these consumer choices the level and mix

“physician-patient pair”, a centaur-like being combining the information possessed by the physician with the circumstances and preferences of the patient. But apart from its inherent implausibility, this attempt to justify ignoring asymmetry of information brings serious theoretical difficulties of its own.

¹¹ The shift of language from “uncertainty” to “risk” is significant. Risk is quantifiable; uncertainty is not. The chance of a well-defined event occurring, say one in ten, or one in one hundred, is a measure of risk. Uncertainty implies not only not knowing what is going to happen, but also being unsure about the nature of the possibilities themselves, let alone their relative probabilities.

¹² Much confusion can be created at this point by introducing the observation that several European systems – the British, for example, or the German -- include a small “upper tier” of private payment by the wealthy for perceived superior quality services. This partitioning of the population is precisely what the trade-off analysis is *not* about. Rather it focuses firmly on the “representative agent”, the hypothetical average individual in the population, and addresses the question of the optimal mix of payment channels for her.

of health care demand. Competitive, for-profit provider firms then respond to these demands in the usual textbook fashion, so that the actual level and pattern of health care utilization – which is measurable – is assumed (subject to the possibility of disequilibrium markets) to be equal to and determined by consumer choices.¹³ To meet these consumer demands, providers must purchase resources or factors of production, and their (derived) demand for these inputs interacts with the factor supply functions of households to generate the corresponding income flows back to those households. The causal sequences in Figure 1 thus all unfold from the upper left, so that is where the key policy levers are and much of the research by mainstream economists has been concentrated.

Apart from its complete “avoid[ance] of distributional considerations” the most immediately obvious silence in this story is with respect to the rest of the financing mix. Health insurance itself is conceived of and analysed as a form of commodity, a financial “product” as private insurers like to call it, supplied by private for-profit firms and purchased by consumers. Yet in high-income OECD countries other than the United States this form of coverage plays an insignificant role in financing health care and even in the United States the public sector contributes (directly and through tax-expenditure subsidies) nearly three dollars for every one coming from private insurers. Analysis of the dynamics of real-world insurance markets readily explains the reasons for the disjunction between academic literature and actual coverage (Evans, 2005). But the actual pattern is from the mainstream economic perspective a major anomaly that cannot be explained, casting doubt on the overall causal story. Why is almost all health insurance public, not private, and what difference does that make for the “physiology” of Figure 1?

But this anomaly is by no means the only one. The right-hand side of Figure 1 attempts to elaborate a set of categories of firms and payment mechanisms (barely) adequate to reflect the complexities of real world systems. Most of these firms, and their sources of funding, bear little relation to the competitive for-profit firms, paid per unit of service, that the economic perspective imports, usually implicitly, from the economic theory textbooks. Why is the health care supply-side organized so differently, in every country, and what difference does *that* make?

The assumption that health care simply comes onto the market at prices reflecting its marginal resource cost, to be purchased by consumers who are then reimbursed in whole or in part by private insurers – a surprisingly common implicit assumption in the academic literature – amounts to assuming that the complexities of health care organization and funding are without point and without effect. Not only should they not exist, but for all practical purposes they do not exist. The production of health care, and the determination of product prices and factor earnings somehow takes place *as if* the industry were made up of for-profit firms operating in perfectly competitive markets, and the resulting patterns of inputs, factor returns, outputs, and prices can therefore be interpreted as they are in standard economic theory, as determined by the tastes of consumers and factor suppliers within the constraints of existing technology and overall resource availability.

¹³ The term “demand”, at least as used by economists, embodies certain behavioural assumptions. It is the quantity of a commodity that consumers are willing to purchase, at given prices and income levels (and states of expectations about future prices and incomes). “Utilization” of health care refers simply to the observation of what was actually used, with no implications as to whose decisions and actions led to the observed pattern. To use the two terms interchangeably is thus to make an implicit assumption of considerable significance. In general, it is safest and least presumptuous to refer to “utilization” unless one has very solid grounds for doing otherwise.

Such an assumption suggests serious perceptual difficulties.¹⁴ At the very least, those relying on it should have to carry a very heavy burden of justification. We would recommend rather that one begin from the presumption that these structural complexities are real, not mirages, they have evolved for good if not necessarily sufficient reasons, and that they have important effects, for both good and ill, on the way health care systems operate.

The naïve clinical perspective provides a ready and powerful answer to the question of why health care provision is organized the way it is. Health care systems are designed to provide people with the care they need, not whatever care they are willing and able to pay for. Different mechanisms serve different objectives. The positive component of the clinical perspective asserts that patients will in fact get the care they need, on three conditions:

- Care is provided by or under the direction of a qualified professional practitioner;
- Sufficient resources of all types are available to meet the needs identified by those practitioners; and
- Patients' access to that care is not impeded by financial or other (language, geography, social distance) barriers.

Give the professionals the tools (and ensure that patients have access to them) and they will finish the job.

This positive claim is rooted in the whole-hearted recognition and acceptance of the asymmetry of information assumed away in the mainstream economic perspective. When it comes to evaluating the impact of health care on health, not just in general but for a particular patient in particular circumstances, there is a presumption that "doctor knows best". If patients are to get the care they need, they must be able to rely on professionals to act for them, recommending or providing the services most likely to improve their health rather than those most desired by the patient or most profitable for the provider. That in turn requires both that the provision of services be restricted to those having the demonstrated competence to identify and provide what is needed – no free entry – and that they be relieved from the pressures faced by firms trying to maximize profits – or simply survive – in a competitive marketplace.

From this perspective, the flows in Figure I unfold from the professional identification of needs, of opportunities to improve the health status of particular individuals by the provision of particular services. This defines the "right" level and mix of servicing, and the corresponding requirement for resources. So long as these are sufficient to "meet all needs" such that no person in this society has a health condition that could be improved by the provision of any form of health care, then professional direction of the health care system, at individual and institutional level, will produce the appropriate care. This requires that provider firms be structured, regulated and funded so as to ensure that decisions about what to do, for whom, how, and when are firmly in professional

¹⁴ One does find more extreme advocates of free markets accepting that the organization and funding of health care provision is indeed peculiar, relative to more "normal" industries, but that it should not be. The differences do matter, but the world would be a better place if providers were all converted to strictly for-profit status in a truly competitive market environment, and relieved of most of the regulatory and self-regulatory apparatus. This position amounts to assuming that current arrangements have no good justification, that they are the result of a great mistake – or a conspiracy – but at least it recognizes that the real world exists.

hands, and that those hands are guided solely by patient needs rather than by provider motives, financial or otherwise. The consumer sovereignty of the economic perspective is replaced by “producer sovereignty” – it is provider decisions that determine levels and patterns of health care use, but those in turn are determined by patient needs.

The clinical perspective is largely silent on the relationship between the structure of the financing system and the appropriate pattern of care provided, except by implication. If patients are to get whatever care they need, then presumably financing must be structured so as to raise sufficient money to pay for the necessary resources, and no one must be denied needed care because of inability to pay. This would seem to imply wholly or predominantly third-party financing. But since private insurance does not and cannot ever cover more than a relatively small part of the costs of modern health care systems, the clinical perspective would seem logically to imply what in fact we observe – an overwhelming predominance of public funding. The clinical perspective thus does provide straight-forward answers, if not spelled out in detail, to the questions: “Why insurance?” and “Why public insurance?” So access to needed care will not be restricted by inability to pay, or insufficient resources.

Yet clinicians themselves have often argued, individually and collectively, *against* public insurance – most notably in the United States. Their resistance is partly rooted in ideological objections that are wholly inconsistent with the clinical perspective as sketched out here. (A surprising number of physicians, particularly in the United States, seem to share the economic perspective – people should *not* get needed care if they cannot afford it.) But there is a more general concern that government financing will never be sufficient to cover all the needs, not because governments cannot pay but because they will not. A political agenda of cost containment will inevitably lead to “underfunding” of health care.¹⁵

At this point, however, we encounter one of the central silences in the clinical perspective. As emphasized above, and in every first course in economics, dollars are not the same as resources. “Meeting needs” requires health care, and providing health care requires the real resources of human time, skills, *etc.* But health expenditures are the product of output *and* price levels, $P \times Q$, and price levels are linked to earnings rates or more generally factor supply prices. A claim of “underfunding” may rest on a genuine perception of unmet health needs, and a corresponding need for an increase in services and resources. But it may equally well be a pay claim, an argument for increases in prices to increase the relative incomes of those working in or otherwise supplying resources for health care. The clinical perspective (in contrast to clinicians themselves) has nothing to say about how rates of reimbursement are or should be determined in the health care sector.

The economic perspective, by contrast, has a perfectly coherent story about how wages and other factor returns are determined in competitive markets, where the purchasers are for-profit firms. In such a context, under the usual stringent assumptions about market structure and participant conduct, well-defined factor supply functions rigidly link the quantities of resources supplied to the prices offered for them. Volumes of services demanded, and the factors inputs necessary to produce them, are all determined interdependently with their prices. Once the volume of services demanded is determined

¹⁵ It might be thought a mark of naiveté to imagine that private insurance could ever finance a significant share of health care expenditures. But the clinical position in practice may be more subtle. *De facto* large covert subsidies to private insurance, as in the United States and Canada, tap public funds without permitting corresponding accountability or control – the best of both worlds if ones’ concern is always to get more money into health care.

by consumer decisions in response to levels of out-of-pocket costs, user charges, then the whole system is determined.

Unfortunately this is a coherent story about an imaginary world. The story goes silent when, as in the health care sector, the purchasers are not competitive for-profit firms, and the resource owners are permitted and even encouraged to limit entry and to collude in price and wage bargaining. Even the major for-profit firms, pharmaceutical marketers, have been permitted to hedge their markets about with patent protection so as to suppress direct price competition. Thus the levels of resources supplied and their rates of reimbursement can and do move independently of each other, at least within a substantial range, and are subjects of intense political bargaining between resource suppliers and reimbursers in public payment systems. It is often in the interests of unions or professional associations – or drug companies – to *claim* that the rigid link between earnings and resource supply postulated in the economic framework actually obtains – higher wages/fees/profits are essential to bring forward enough nurses/physicians/new drugs to “meet needs”. But in the health care sector in particular, even when the claim has some validity there is always a good deal more to the story.

There is also more to the story of price determination than wage and other factor price levels. Economists conceptualize the relationship between resource inputs and commodity outputs as a “production function”. This specifies the maximum level of output that can be produced with currently available technology from any given combination of inputs. But the production function specifies a frontier, a boundary to the “feasible set” – it is always possible to produce less than available resources permit, or equivalently waste resources in producing the current level of output.

The boundary is reached, at least in theory, by a profit-maximizing firm. In order to maximize profits, the firm must produce its outputs at minimum possible cost; this in turn implies using no more resource inputs than necessary.¹⁶ Firms in the real world will conform more or less closely to this theoretical ideal depending upon the intensity of competition in product or factor markets. But there is no basis whatever for assuming that the equality holds in firms organized for purposes other than profit maximization, and/or exempt from product market competition or capital market takeover.

The clinical perspective is completely silent as to what other mechanisms might operate to assure technical efficiency in the health care sector. Yet technical efficiency is implicit in all claims of “underfunding”: service output cannot be raised without acquiring (and paying for) more resources. Conversely pressures for “hospital downsizing”, or more generally requests to “do more with less”, assume implicitly that for much of health care production there *is* at current resource levels a substantial discrepancy between what is and what could be produced.

There is however another possibility. A given bundle of health care resources may fall short of its maximum potential for improving population health either because the production of health care is taking place at less than full technical efficiency, or because the care that *is* being produced is not the most appropriate to meet patients’ needs. The positive component of the clinical perspective amounts to asserting that, under the conditions above, the care patients get is the care they need – otherwise it would not be

¹⁶ These are necessary, not sufficient, conditions for profit maximization. The profit-maximizing firm must not only minimize the cost of its output, but also select the right level of output. And cost minimization requires both not wasting resources, and choosing the least-cost combination of resources, given their relative prices.

provided – but is silent as to the technical efficiency with which that care is produced. In practice cost containment efforts focus on both “production functions.”¹⁷

The Eclectic Structuralist perspective, as noted above, includes essentially the same normative position as the clinical, people *should* get the care they need, the care that improves their health, more or less independently of their ability to pay. But it takes dead aim at the central *positive* presumption of the naïve clinical perspective, *i.e.* that, under appropriate professional direction, they do. Holders of this perspective have assembled increasingly large and sophisticated batteries of fact and argument to show that the care patients actually receive depends upon much more than evidence of potential capacity to benefit. (Most prominent among this school have been the researchers assembled by John Wennberg and Elliott Fisher at the Dartmouth Medical School. Fisher (2007) provides a comprehensive and very accessible compilation of their findings.)

Structuralists certainly do not claim that there is *no* connection between patient needs and health care use – demonstrably false and patently absurd. But this perspective focuses attention on the powerful effects of other factors influencing patterns of use by individuals and particularly by populations. Recalling Rose’s Law (Rose, 2001), that the causes of rates are not the same as the causes of cases, relative needs may be the primary, indeed the overwhelmingly dominant, factor determining who gets what care *within* a population, and yet play little or no role in determining comparative rates of care *between* populations, or in the same population over time.

How needs are defined and recognized, and especially how and how extensively they are responded to, are highly variable both within and between health care systems, and shift over time, for reasons that have no detectable connection with patient conditions or outcomes. The literature on “clinical variations” has been accumulating for over forty years, showing highly variable patterns of practice among different practitioners, institutions, regions, and countries. This literature has traditionally been dismissed with, in effect, the comment “Who knows which rate is right?” and thus ignored.

In the last decade, however, increasingly sophisticated analysis of very large data sets has shown that at least in the United States, regions and institutions with higher rates of intervention and cost do not show better outcomes for patients, in either health or satisfaction, and indeed, in aggregate, more can be harmful. The most prominent research program has been the work of researchers assembled by John Wennberg and Elliott Fisher at the Dartmouth Medical School. Fisher (2007) provides a comprehensive and very accessible compilation of their findings.

This is a fundamental challenge to the positive component for the naïve clinical perspective; so far the collective response of health care systems has been simply to hope that these findings will just go away. “*Underfunding* is the issue; never mind what we do with the money!”

The Structuralist perspective focuses attention on the different forms and amounts of capacity in the health care system, and on the incentives, in particular but not exclusively economic, embodied in the way that that system is organized and financed. Within Figure I, therefore, it emphasizes the importance of the right-hand side – on what terms does money flow to the provider organizations, and what are the motivations of

¹⁷ The reference to hospital downsizing – the dramatic decline in in-patient utilization in a number of countries over the last quarter of the twentieth century, flags an inherent ambiguity in the definition of “output” or the commodities produced by health care providers – particularly hospitals. Do hospitals produce in-patient days, or treated cases, or simply a wide array of particular services and procedures? Is an unnecessarily long length of stay for a clearly necessary surgical procedure, the inefficient production of the commodity “a treated case” or overprovision of the commodities “in-patient days”?

those organizations? A for-profit corporation with the objective of maximizing “shareholder value” (profit) will behave differently from a not-for-profit hospital or the private practice of one or more physician partners – and they do.¹⁸

It may seem obvious that the behaviour of an individual or an organization will be different, depending upon how that individual/organization is paid and what s/he/it is trying to do – and to Structuralists it is. But acceptance of that “obvious” point implies rejection of the *a priori* claims that, in practice, health care utilization patterns are solely determined either by clinicians’ judgements of need, or by patients’ willingness/ability to pay. There remain important questions as to the relative strengths of these different factors in influencing levels and patterns of health care utilization, but those are empirical questions.¹⁹

The classic example of the Structuralist perspective is “Roemer’s Law” of hospitals, that “a built bed is a filled bed”. Roemer (1961) observed, from a natural experiment, a strong correlation between the number of beds available in a particular region, and the rates of hospital utilization. He inferred that this represented a causal connection. Others argued that a more generally observed correlation between capacity and use might reflect a reverse causality, that areas of high need might build more beds, but subsequent studies confirmed Roemer’s interpretation. The primary causal mechanism may operate through the influence of “time and trouble costs” for physicians seeking to admit patients, rather than an absolute capacity constraint. Hospitals typically run at occupancy rates well below 100%, and more bed capacity (beds per capita, age-adjusted) tends to be associated with lower occupancy, but also with increased use rates for the relevant population.

Subsequent declines in the use of hospitals have, however, underlined an implicit qualification to this “Law” – a built *and reimbursed* bed is a filled bed. The introduction of case-based reimbursement in the American Medicare program in 1988, such that the amount of reimbursement a hospital received (under Part A for inpatient care) for a patient with a particular diagnosis was based on a predetermined schedule and no longer linked to actual costs incurred, was associated with an immediate and continuing decline in hospital lengths of stay, and a transfer of diagnostic and other procedures to free-standing (and separately reimbursed) facilities.

Fifty years later, the studies of the Dartmouth program (Fisher, 2007) have confirmed Roemer’s insight – roughly half of the regional variation in use among the American Medicare population is associated with variation in the availability of facilities and personnel. The remainder seems to be linked to clinical habits and cultures. None of the regional variations could be associated with patient needs or preferences.

¹⁸ It is hard to imagine a private practitioner amassing the string of criminal convictions and billion-dollar fines assembled by Pfizer and treated as simply a business expense (Evans, 2010).

¹⁹ The answers can be disturbing. For example, nearly a decade ago American spinal surgeon Edward Benzel conjectured that probably fewer than half of the spinal fusion procedures then performed in the United States were appropriate (Abelson and Petersen, 2003). This is very serious surgery (and very expensive). Yet “The reality is, we all cave in to market and economic forces.” (spinal fusion is also very lucrative). The reimbursement system is “totally perverse”.

Summary:

The central concern of all three alternative perspectives, or paradigms, is the determinations of health care utilization. For the naïve clinicians, utilization *should* be determined by patients' needs, as judged by clinicians. And it *will* be, subject to the availability of appropriate resources, professionally directed, and the removal of access barriers. This paradigm has very powerful intuitive appeal, to clinicians and patients alike. Research focuses on identifying presently unmet and new needs, and finding better ways to meet them.

For the mainstream economist, utilization *should* be determined by patients' preferences and personal resources, as expressed in their willingness to pay at prices reflecting resource opportunity costs. And it will be approximated, subject to the trade-off imposed by insurance that reduces risk but promotes and "overuse". The paradigm has minimal normative appeal or positive plausibility even to economists outside their professional work. Its strength derives from the fact that its predictions and prescriptions can be rigorously formulated, and that it yields recommendations with strong and highly regressive distributive implications. Research focuses on measuring a hypothetical "elasticity of demand" so as to optimize a hypothetical "trade-off" between risk and an idiosyncratic concept of "overuse". Holders of this perspective seem ambivalent as to whether this would require dismantling the extensive regulatory structure erected in response to the clinical perspective, or whether that structure is simply a mirage that can be ignored.

Eclectic Structuralists essentially accept the normative position of clinicians, but reject their positive claims on empirical grounds. Extensive analysis of actual utilization patterns shows very powerful effects from capacity levels, and the incentives faced by and the motivations embedded in provider organizations. Research focuses on provider behaviour in differing settings, and its relation to evidence of the effectiveness of care.

* Correspondence to: Robert G. Evans, International Health Economics Association.
E-mail: bevans@chspr.ubc.ca

References

- Abelson, R., and M. Petersen (2003) "An Operation to Ease Back Pain Bolsters the Bottom Line, Too" *New York Times*, December 31.
- Arrow, K.J. (1976) "Welfare Analysis of Changes in Health Coinsurance Rates" in R. Rosett, ed. *The Role of Health Insurance in the Health Services Sector* New York: National Bureau of Economic Research, pp. 3-23.
- Culyer, A.J., and R.G. Evans (1996) "Mark Pauly on Welfare Economics: Normative Rabbits from Positive Hats" *Journal of Health Economics*, Vol. 15, no. 2 (May) pp. 243-51.
- Evans, R.G., (2005) "Preserving Privilege, Promoting Profit: The Payoffs from Private Health Insurance" in C. M. Flood, K. Roach and L. Sossin, eds. *Access to Care, Access to Justice* Toronto, University of Toronto Press, pp. 347-368.
- Evans, R.G., (2010) "Tough on Crime? Pfizer and the CIHR" [Commentary] *Healthcare Policy* Vol. 5, no. 4, 2010, pp. 16-25.
- Feldstein, M.S., (1973) "The Welfare Loss of Excess Health Insurance" *Journal of Political Economy* Vol. 81, no. 2 Part 1 (March-April) pp. 251-280.
- Fisher, E.S. (2007) Pay-for-performance: More than rearranging the deck chairs? Robert and Alma Moreton Lecture, Dartmouth Medical School, May 21.
- Fuchs, V.F. (1966) 'Economics, values and health care reform' *American Economic Review* Vol. 86, no. 1 (March) pp. 1-24.
- Pauly, M.V. (1969) "A Measure of the Welfare Cost of Health Insurance" *Health Services Research* Vol. 4, no. 4 (Winter) pp. 281-292.
- Pauly, M.V. (1996) "Reply to Anthony J. Culyer and Robert G. Evans" *Journal of Health Economics* Vol. 15, no. 2, pp. 253-4.
- Reinhardt, U.E. (1992) "Reflections on the Meaning of Efficiency: Can Efficiency Be Separated from Equity?" *Yale Law & Policy Review* Vol. 10, no. 2, pp. 302-315.
- Reinhardt, U.E. (1997) "Abstracting from Distributional Effects, this Policy is Efficient., in Barer M.L., T.E. Getzen and G.L. Stoddart, eds. *Health, Health Care and Health Economics*, New York: Wiley & Sons: 1-53.
- Roemer, M.I. (1961) "Bed Supply and Utilization: A Natural Experiment" *Hospitals: Journal of the American Hospital Association* Vol. 35: pp. 35-42.
- Rose, G. (2001) "Sick individuals and sick populations" *International Journal of Epidemiology* Vol. 30, no. 3, pp. 427-432.
- Roos, N.P., E. Forget, R. Wall and L. MacWilliam (2004) "Does universal comprehensive insurance encourage unnecessary use? Evidence from Manitoba says 'no'" *Canadian Medical Association Journal* Vol. 170, no. 2 (January 20) pp. 209-214.
- van Doorslaer, E., A. Wagstaff, and F. Rutten, eds. (1993) *Equity in the Finance and Delivery of Health Care: An International Perspective* New York: Oxford University Press.
- Woolhandler, S. and D.U. Himmelstein (2002) "Paying for National Health Insurance – and Not Getting It" *Health Affairs* Vol. 21, no. 4 (July-August) pp. 88-98.