User Charges, Snares and Delusions:
Another Look at the Literature

Greg L. Stoddart
Morris L. Barer
Robert G. Evans

HPRU 93:14D
December 1993
User Charges, Snares and Delusions: Another Look at the Literature

Greg L. Stoddart
Morris L. Barer
Robert G. Evans

HPRU 93:14D December 1993
The Centre for Health Services and Policy Research was established by the Board of Governors of the University of British Columbia in December 1990. It was officially opened in July 1991. The Centre's primary objective is to co-ordinate, facilitate, and undertake multidisciplinary research in the areas of health policy, health services research, population health, and health human resources. It brings together researchers in a variety of disciplines who are committed to a multidisciplinary approach to research, and to promoting wide dissemination and discussion of research results, in these areas. The Centre aims to contribute to the improvement of population health by being responsive to the research needs of those responsible for health policy. To this end, it provides a research resource for graduate students; develops and facilitates access to health and health care databases; sponsors seminars, workshops, conferences and policy consultations; and distributes Discussion papers, Research Reports and publication reprints resulting from the research programs of Centre faculty.

The Centre's Health Policy Research Unit Discussion Paper series provides a vehicle for the circulation of preliminary (pre-publication) work of Centre Faculty and associates. It is intended to promote discussion and to elicit comments and suggestions that might be incorporated within the work prior to publication. While the Centre prints and distributes these papers for this purpose, the views in the papers are those of the author(s).

A complete list of available Health Policy Research Unit Discussion Papers and Reprints, along with an address to which requests for copies should be sent, appears at the back of each paper.
Preface

This is one in a series of articles by the authors about the ongoing debate over user charges in the Canadian health care system.

In this paper we review and extend an earlier, in-depth analysis of the effects of user charges which three of us (MLB, RGE, GLS) published in 1979. The present paper assesses whether experience and published literature in the years since then alter any of the (largely negative) conclusions of the earlier study concerning the ability of direct charges to patients to achieve important public policy objectives, including controlling health care costs.

Other papers in this series discuss frequently heard arguments for user charges, and focus on specific, and sometimes more technical, dimensions of the user charge debate. A brief description of each paper follows.

"Why Not User Charges? The Real Issues", describes and analyzes the most frequently heard arguments for user charges and the evidence that exists for claims and counter-claims that are often made. We explore the arguments carefully, asking what they really mean, what values they are based on, and what fundamental issues are at the heart of the user charge controversy.

"The Remarkable Tenacity of User Charges" concisely documents the history of the user charge debate in Canada. It reviews the participation, positions and rationales of Canadian interest groups in debates over patient participation in health care financing.

"Who Are the Zombie Masters, and What Do They Want?" focuses on those people and organizations who have consistently revived and promoted the idea that user charges will help meet a number of important social policy objectives, despite the fact that such charges have been repeatedly rejected by policy-makers and the general public (and the claims of their supporters refuted by analyses of the effects of such charges). We identify a number of distinct groups of "zombie-masters" and find that they seem motivated largely by the expectation that they, or the people they represent, will benefit in some way from the (re-) introduction of user charges.

"Charging Peter to Pay Paul: Accounting for the Financial Effects of User Charges" outlines a formal and comprehensive analytic framework in which income transfers - the principal effects of user charges - can be traced between groups in the population (e.g. the healthy, the sick, the rich and the poor), between payers and health care providers, and among providers. The paper uses the framework to analyze the income transfers associated with different types of user charges.

"It’s Not the Money, It’s the Principle" examines why user charges exist for some health care services and not for others. The paper analyzes the characteristics of services which (do or should) underlie decisions to charge in part or in whole for specific types of services.

In addition, a bibliography entitled "User Charges in Health Care" provides an extensive set of references to articles of relevance to the user charge debate in Canada, drawn from diverse sources including academic research and policy analysis literature, the popular press, government
documents and reports, and the publications and reports of non-governmental organizations including the professional associations representing a variety of health care providers.
User Charges, Snares and Delusions: 
Another Look at the Literature

Greg L. Stoddart  
Centre for Health Economics and Policy Analysis, and  
Department of Clinical Epidemiology and Biostatistics  
McMaster University

Morris L. Barer  
Centre for Health Services and Policy Research, and  
Department of Health Care and Epidemiology  
University of British Columbia

Robert G. Evans  
Department of Economics, and  
Centre for Health Services and Policy Research  
University of British Columbia

December 1993

This work was funded by the Ontario Premier's Council on Health, Well-Being and Social Justice. The research assistance of Vandna Bhatia is gratefully acknowledged. The authors wish to thank the many individuals both inside and outside the health care system who have taken the time to share their views on user charges. Responsibility for the views expressed herein, and any errors or omissions, rests solely with the authors. R.G. Evans is supported by a National Health Scientist Award from Health and Welfare Canada, and is a Fellow of the Canadian Institute for Advanced Research. G.L. Stoddart and M.L. Barer are a Fellow and an Associate, respectively, of the Institute.
1. Introduction

In 1978, following the release of the Taylor Report [1] recommending the limited reintroduction of direct charges to patients for health care in Ontario, we were commissioned by the (now defunct) Ontario Economic Council to examine the likely effects of such user charges, with particular emphasis on their ability to control health care expenditures. The results of our detailed analysis of several then popular proposals for "patient participation in financing" were published the following year by the Council in the monograph Controlling Health Care Costs by Direct Charges to Patients: Snare or Delusion? (hereafter referred to as OEC79) [2].

Although the title may have struck some readers as unnecessarily flippant, its connotations captured well the conclusions of our analysis, as illustrated by the following three paragraphs drawn from the preface to the monograph:

Our analysis indicates that direct charges will benefit providers, private insurance companies, and the provincial government. Direct charges in most forms (particularly the ones most frequently recommended) will serve as an injection of additional funds into the sector and thus as a source of increases in provider incomes. Furthermore, exposure to any significant direct charge is likely to lead consumers to seek supplementary private insurance coverage. Finally, direct charges provide a means of keeping the lid on health care expenditures in government budgets while allowing total (public plus private) expenditures to rise. The snare is not likely to end up empty. It will be filled by those unfortunate enough to become ill, for it is they, and only they, who will feel the effects of direct charges.

The delusion is embodied in the recurring argument that one way to solve the 'health care cost crisis' is to make patients more aware of the cost of providing their care. The argument has two prongs. First, it is asserted that a significant amount of patient-initiated utilization consists of 'unnecessary' care which could be deterred by direct charges. Second, it is often suggested that individuals should assume greater personal responsibility for their health status through preventive lifestyle changes and should rely less on the medical profession, a transition which would be helped along by direct charges.

Upon further examination, however, we find the argument rests on very shaky ground. There is little, if any, evidence to suggest that patients are the primary generators of marginally needed care and (perhaps because of that) no evidence whatsoever to suggest that prices tend to deter that segment of care first. Therefore, while there are a number of potential avenues for introducing personal accountability, analyses of them converge upon the same conclusions - consumption of necessary care may be deterred, aggregate health care expenditures are influenced marginally, if at all, and there is little reason to believe that direct charges for health services encourage preventive self-maintenance. Moreover, deterrence of care-seeking by an individual does not in itself reduce health care use, since additional provider-generated utilization can easily offset this reduction.

[2,p.viii]

Although we identified some limited and very selective potential roles for direct charges
to patients, our analysis was generally critical of the direct charge concept and its supporters. Despite the resilience of proposals for user charges in Canada and their widespread use in the United States, we characterized the case against most forms of such charges as "surprisingly strong". We concluded the preface by stating, "It appears that the direct charge concept is indeed an idea whose time has gone." [2,p.x]

This has turned out not to be the case, however. Despite the elimination of 'official' extra-billing across Canada following the passage of the Canada Health Act in 1984, the calls for some form of user charge for health care have persisted. Recently they have escalated as revenue-hungry provincial governments examine their financing options amidst a lingering recession, a continued withdrawal of federal contributions for health care, and steady pressure from health care providers for higher expenditures and incomes.

Thus too have been drawn back into a re-examination of user charges for health care in today's context. In other papers that comprise our current project, we document the evolution of proposals for user charges by various interest groups in Canada [3] and analyze the reasons for the resilience of the idea [4]. We also examine in detail the popular arguments for charges [5] and emphasize their redistributive effects on access to care, the burden of paying for care, and health outcomes [6,7].

In this paper, we examine whether the research literature of the past fourteen years changes any of the conclusions drawn in OEC79. The first half of the paper reviews and summarizes the original monograph, including its objectives, analytic framework, synopsized arguments and evidence, and conclusions. The second half of the paper provides a review of relevant literature since 1979 and an assessment of its impact on the original analysis. The literature review focuses primarily on Canada and the United States, but places the North American experience in an international context. It is not intended to be either detailed or exhaustive, although we have attempted to find and incorporate the major relevant studies. As was the case for OEC79, the focus is also primarily (though not exclusively) on hospital and physician services. Elsewhere we discuss the rationales for user charges on other types of services [8].

2. The OEC79 Analysis
2.1 Objectives and Background

The objective of the study was to examine the relationship between health care expenditures and direct charges to patients, drawing on analyses and available empirical evidence about both the nature of health care 'markets' (including the behaviour of both patients and health care providers) and the effects of specific forms of direct charges. The study examined various ways in which direct charges to users of health care services might be structured, and how each variant might influence total health care expenditures as well as other objectives of the public hospital and medical insurance programs in Ontario.

Total health care expenditure - i.e. the sum of public and private expenditures, rather than just public expenditure - was chosen as the focus of the study because:
Arguments for 'privatization' and 'control of the size of the government sector', which recommend shifting part of the burden of health costs from public budgets to private individuals as patients or payers of private insurance premiums, need careful scrutiny to ensure that they are not in fact proposals which increase total health costs while reducing the government’s share... [2,p.3, emphasis added]

Whether health care expenditures are financed through direct taxes (e.g. the personal income tax), public or private insurance premiums, or direct charges to patients, it is ultimately Canadian citizens who bear the cost of health care, in one form or another [6]. Although direct charges to patients might (and even this is far from certain) lead to a reduction in the government share of health care spending, such a reduction would at best be cosmetic if in fact Canadian citizens found themselves paying higher costs overall, when their private payments were added onto their other, public contributions. It is the total share of national or personal income devoted to health care expenditure, and therefore not available for other things, that is - or should be - the focus of policies aimed at controlling health care 'costs'.

Of course, not every Canadian would be worse off. Depending on the amount of taxes they paid, the level of the direct charges, and their use of health care, some Canadians would, individually, be better off while some would be worse off as a result of a direct charge policy [5,6]. This is as true today as it was in 1979. The redistributive effects of direct charges on the burden of paying for care were included in OEC79 and are discussed below.

Premiums were not considered to be a direct charge because the amount an individual or family paid (both then and now, in the provinces where premiums still exist) was not in any way linked to actual or expected use of health care services. Direct charges (which are typically referred to today as 'user charges') are by definition a function of use. The specific functional form may vary considerably, as shown in the next section, but the characteristic common to all variants of direct or user charges is that the more care an individual uses, the more the individual pays out-of-pocket, at least up to some cutoff or maximum level and for those individuals not exempted from paying the charges.

The health policy context in which the original monograph was written warrants a final introductory comment, because its eerie similarity to the policy environment today vividly demonstrates how difficult and frustrating serious reform of the health care sector really is. Key elements of the environment identified in the background chapter of the monograph [2, chpt. 1] were:

1) public perception of a health care cost 'crisis' attributable to universal, public,
first-dollar insurance coverage for hospital and medical services, despite empirical
evidence to the contrary,

2) no significant changes to the supply side of the hospital and medical industry to
accompany the almost total displacement of the market by public payment on the
demand side, despite "a continuing series of studies diagnosing and documenting
the essentially unmanaged nature of health care delivery " (p.5),

3) the slowing of growth in the economy generally and the increasing public concern
over the size and scope of governments, which in turn increased the perceived
political risks to raising taxes,

4) a steady build-up of political pressure from physicians and hospital administrators
to relax constraints on their incomes and budgets,

5) numerous inquiries and visits by American health professionals and policy-makers
(in the wake of failed discussions about a national health insurance initiative south
of the border), which "guaranteed continued discussion among their Canadian
counterparts regarding deficiencies in and disappointments with several aspects of
the Canadian system" (p.8), and

6) growing concern and dissatisfaction with the level of population health that the
level of health care expenditure was ‘buying’, and an increasing uneasiness
(following the Lalonde Report [9] and the work of McKeown [10]) that increasing
resource allocations to health care was not the best way to improve the health of
Canadians.

This listing is so remarkably similar to today’s environment that it is hard to believe it
was written almost fifteen years ago. Indeed, it is hard to find a better concise summary of the
context in which we are conducting our current project!

2.2 Analytic Framework

The analytic framework had two components - a set of policy objectives on which the
impact of direct charges was evaluated, and a taxonomy of different types of direct charges.

Policy objectives were divided into two categories labelled "health insurance objectives"
and "health care delivery system objectives". The health insurance objectives, risk reduction and
wealth transfer, reflected the two main reasons why Canada had established a universal,
compulsory, publicly-financed insurance plan for (first) hospital and (later) medical care, instead
of opting to continue with or revise a system relying on private insurers.

The first goal was to reduce the risk of unexpected and unpredictable financial loss due
to illness, and to do so for all Canadians, including high-risk and very ill individuals whom
private insurers typically exclude from coverage, either by refusing coverage or by pricing
coverage beyond what they can afford. The second, related goal of the public plan was to provide access to care based on need, irrespective of ability-to-pay, and to redistribute the overall cost of the health insurance plan through public financing (relying heavily on the personal income tax) largely on the basis of ability-to-pay. By making the receipt of needed care independent of an individual's ability-to-pay for it, the architects of the Canadian system embodied a wealth transfer function in it. Wealth is transferred from individuals with low use of services to those with high use, other things being equal.

This set of social values reflected the wishes of Canadians before and during the establishment of public health insurance and still commands broad political and public support today, not only in Canada but also internationally [11].

Health care delivery system objectives focused on the issues of utilization and costs, and embodied objectives of both technical and allocative efficiency. The term "technical efficiency" referred to the objective of producing necessary health care services with minimum resource commitments, i.e. at least cost; wasted resources are a loss to society (in public or private systems). But society also has an interest in seeing that the amount and types of services that are provided and utilized are appropriate to the needs being addressed and the specific clinical circumstances involved. Inefficacious or ineffective services, no matter how efficiently they are produced, also waste resources if they do not improve the health of patients. Therefore, policy

---

2 "One may not find decreases in the expected net wealth position of all individuals with low probability of use since, depending on the progressivity of the tax system, some expected low users might also pay very little into the plan. Similarly, under a progressive tax system there may be little change in the expected net wealth position for high-income/high-use individuals. In this situation, the wealth transfer would be from low users with high income to high users with low income, rather than simply from low to high users." [2, p.21, note 2]. For a more detailed treatment of this, see Evans, Barer and Stoddart [6].

3 "Minimum resource commitments" may be interpreted in either a real or nominal context. The former refers to minimizing the amounts of physical inputs, e.g. physician time, hospital beds, drugs, etc. used to produce a given level of service. The latter refers to minimizing the dollar cost of a given level of service, and takes into account the prices of the inputs. It is sometimes distinguished from the physical definition of efficiency by the term "cost-effectiveness" [12], although this term was not used in OEC79.
objectives also included a concern for appropriate levels and patterns of utilization.\(^4\)

The taxonomy of different types of direct charges was based on two distinctions:
1) whether the charge was uniform or varied across different providers, and
2) whether the charge was determined by providers (individual practitioners, institutions or their associations) or by third parties such as governments or insuring agencies.

The distinction between uniform and differential pricing was felt to be important because it related directly to the scope for attaining technical efficiency in production. (This is discussed further in section 2.3.5 below.) The identity of the party determining the charge was felt to be significant because the pattern of differential charges which emerges if providers set their own rates can be quite different from that of a charge structure set by third parties and can serve different purposes. For example, charges set by providers themselves, instead of by a third party, may serve primarily to increase providers’ incomes rather than to improve the appropriateness of levels and patterns of utilization.

The specific types of direct charges which were analyzed, and their place in the taxonomy, are shown in Figure 1. (Each type is described in turn as it appears in the synopses below.)

In the OEC79 study we analyzed the likely impact of each type of direct charge on the five policy objectives (risk reduction, wealth transfer, technical efficiency, and levels and patterns of utilization). In each case we proceeded by specifying the assumptions about the health care industry that were necessary to predict the impact of the charge and then drew on existing information to assess the plausibility and consistency of those assumptions and thereby the likelihood that the charge would contribute to the achievement of each of the policy objectives.

\(^4\) In the present climate of concern about the scope of government involvement in society, it bears re-emphasizing that concerns about efficiency are independent of the type of system, public or private, which may exist:

The social interest in health care patterns and levels may be expressed through government institutions and interventions, but it does not arise as a result of government’s increased intervention in health care. On the contrary, the direction of causation is exactly the reverse. Government’s role in health care delivery has grown in response to our interest, as members of a common society, in the health of all of us; it reflects our interdependence. [2,p.23]
Synopses of the key analytic points for each type of charge are presented in the following section, moving clockwise from the upper left quadrant of Figure 1. The section concludes with an overview of the impact matrix for all types of charges and policy objectives.

2.3 Analysis of Specific Types of Direct Charges

2.3.1 Coinsurance

Coinsurance is a form of direct charge in which the patient is required to pay a fixed percentage (say, for example 20%) of all costs incurred on his or her behalf. The baseline analysis assumed a scenario in which fees received by providers from the provincial insurance plans remained constant and patients were billed periodically by Ministries of Health for 20% of all their hospital and medical costs.

If nothing else changes, this simply lowers the cost of health care in the public budget by 20%. If there are no consumer, provider or institutional responses, then the impacts on both risk reduction and wealth transfer objectives are perverse when judged against the intent of the health insurance objectives discussed above and their underlying social values. Significant financial risk is reintroduced for patients with high levels of utilization, and wealth is transferred from citizens-as-patients to citizens-as-taxpayers, a transfer of dubious distributional merit in this context because tax burdens are spread across the population in proportion to income, while illness and its costs are inversely related to income. Coinsurance which is uniform across providers is also devoid of any incentive for providers to improve production efficiency.

Most advocates of coinsurance (and other forms of direct charge), however, typically expect the charge to reduce rather than simply shift the cost of health care. For that to occur, four major assumptions would have to be satisfied:

1) patients must be sensitive to prices in making decisions about health care use,
2) private insurance must not step into the gap to insulate patients against potential out-of-pocket costs,
3) health care providers must not react to decreases in patient-initiated utilization in ways that would offset the initial effect, and
4) patients must pay their bills.

The analysis proceeded to assess the plausibility of these assumptions, although not without some difficulty\(^5\), for both hospital and medical services. The overall conclusion was that

\(^5\) "The problem is that all these behavioural responses are interdependent ... it is extremely difficult to partition differences into patient-generated and provider-generated effects. The pattern of exclusion affects the size of utilization response; if the aged and low-income groups are excused direct charges, the average utilization response is less, because these groups appear most price-sensitive. Who purchases supplementary insurance, and how much, will depend on the magnitude of the direct
it was unlikely that any of the assumptions (with the possible exception of the fourth one) would be satisfied and, therefore, extremely unlikely that the complete set would be satisfied. In addition, the review of available evidence suggested that there was no reason to expect efficiency improvements from coinsurance, and that the actual impact on the remaining objectives was quite likely to be perverse.6

Analysis of the Canadian experience before and after the introduction of both private and public hospital insurance, and analysis of the 1968-1971 Saskatchewan experience with (per service) direct charges of $2.50 per hospital day (about $10 in today’s prices) indicated little, if any, price sensitivity for hospital care. American studies of the effect of out-of-pocket charges on hospital use showed some price sensitivity but were much more difficult to interpret due to a variety of methodologic problems which affected the empirical estimates.

In both cases, price sensitivity might be expected to increase with the amount of the charge, but then the second and fourth assumptions above would be less likely to be satisfied. "Out-of-pocket costs of, say, $100 a day would probably affect use, but after the exclusion of the poor and the aged, and the re-emergence of private insurance, no one (or almost no one) will ever pay such charges!" [2,p.44]. In other words, if charges were set at a level high enough to have an impact, it would not likely be politically feasible to maintain the ban on private insurance, or to fail to exempt the poor and elderly - the groups who use the most care - from payment. Both scenarios would insulate individuals against the effects of charges and dilute an already weak response of hospital use to direct charges. And if charges did not affect hospital use (which at that time made up almost two-thirds of insured health care costs in Ontario), then their effect on total utilization and cost could hardly be expected to be significant.

The evidence on price sensitivity for medical care use was less clear-cut. Studies of Quebec’s experience with the introduction of Medicare indicated no overall increase in utilization in response to the removal of price barriers to utilization, but this is not necessarily the same behavioural response as would occur for the imposition of prices. Analysis of the 1968-1971 charge. Moreover, data from Canadian experience before universal hospital or medical coverage arrived may not adequately represent how people would respond to direct charges introduced into established universal systems." [2,p.29-30]

6 It is not possible here to provide detailed reviews of the evidence or references without essentially reproducing the original monograph. For coinsurance, this material can be found on p. 27-51 of OEC79 [2]. Hereafter, the pagination for the detailed analysis and references is provided in a note accompanying the introduction of each specific form of direct charge.
Saskatchewan experience with per service charges for physician services of $1.50 per office visit and $2.00 per home or emergency visit (about $6 and $8 in today's prices, respectively) indicated some price sensitivity overall (a 5.7% reduction in use), but it was almost exclusively accounted for by a stronger response (about 18% reduction in use) among elderly and low-income groups. (For middle-income groups, and for certain services such as comprehensive examinations that paid the physicians relatively highly, utilization actually increased.)

American studies of the extent to which full first-dollar medical insurance instead of partial coverage could lead to increased utilization of physicians' services (the inverse of the imposition of deterrent fees) at the time provided widely divergent results. They often indicated higher price sensitivity than the Canadian studies, but were plagued by methodologic issues which again rendered the estimates difficult to interpret and of questionable applicability to the Canadian setting.

Foremost among these issues was the failure of the American studies to address the effects of charges on the total utilization of services for a broad population group, taking into account the reactions of physicians to decreases in demand resulting from patient price sensitivity. However, even with these restrictions, our reworking of the most reliable American figures in a hypothetical Canadian context did not yield estimates of price sensitivity significantly higher than those from the (overall) Saskatchewan experience.

The reviews of Canadian and American studies and experience in OEC79 also found that:

a) there was no evidence to suggest that the selective utilization reductions resulting from coinsurance were concentrated on those involving frivolous use or unnecessary services,

b) there was some evidence to suggest that charges discouraged the use of preventive services, and so could potentially affect the health of patients adversely,

c) there was reason, and some evidence, to expect the reactions of health care providers (either generating more utilization, or bargaining for higher fees to replace lost income, or both) to undermine the cost-reducing effects of direct charges, and

d) a government-administered and collected coinsurance plan was likely operationally implausible (and would likely be less effective in reducing utilization than charges levied at the point-of-service), yet the alternative - a scheme in which physicians collected and kept the coinsurance charge - was fraught with even more difficulties.

2.3.2 Deductibles

With this form of direct charge, the patient is required to pay 100% of all bills in a given period up to some maximum amount (the deductible), beyond which insurance coverage takes...
effect. (Deductibles may be combined with coinsurance, yielding a corresponding combination of their effects [8].)

Purely from an insurance viewpoint, and ignoring effects on objectives other than risk reduction, the deductible approach has potential value for services where utilization is relatively common and costs for most users are low. Risk pooling through insurance costs money; risk reduction is the benefit. The latter may not be worth the former to consumers for small and relatively predictable expenses. The deductible serves to divide the population into two groups, one with and the other without risk of high expenditure levels, and focuses the insurance function on the very small part of the population with a high risk of expenditures. While prescription drugs might (for most people) be a type of health care that meets the conditions for a deductible to be useful when judged solely on insurance principles, medical and hospital care are types which do not meet the conditions.

The impact of the deductible form of direct charge on the five policy objectives is identical to that of coinsurance (for levels of expenditure below the deductible ceiling), for the same reasons as in the previous synopsis. The OEC79 analysis concluded that,

A deductible applied to hospital care, for example, would amount principally to a tax which transferred funds in the amount of the deductible from each person admitted to the hospital to the general provincial revenues. Merely from observing the very pronounced age-related pattern of hospital use one would expect such a tax to be highly regressive, shifting wealth from lower- to higher-income groups. [2,p.54]

2.3.3 Per-Service Charges

Per-service charges, as their name implies, are flat-rate charges on specific services, for example the $1.50 "deterrent fee" per physician office visit or the $2.50 charge per hospital day in effect in Saskatchewan from 1968 to 1971.

Again, compared to a universal, first-dollar coverage insurance plan, the impact of per-service charges on the policy objectives was found to be the same as that of coinsurance and deductibles. They would expose patients to increased financial risk, create perverse wealth transfers, and have little or no impact on efficiency. Their impact on demand was expected to be limited and heavily concentrated on lower-income users, with possibly harmful health effects.

Two circumstances were identified in which per-service charges might have merit. The first was for the residential component of chronic care or long-term care, where they would offset self-maintenance expenses saved by patients living in these institutions. The second was in cases where health services had been shown to have little or no health benefit, although the analysis went on to point out that if this truly was the case, then presumably deinsurance (a per-service

---

8 For detailed analysis and references, see OEC79 [2], p. 56-60.
charge equal to the insured fee for the procedure) was warranted.

2.3.4 Income- and Income Tax-Linked Options

The OEC79 analyses of these variants of direct charges are of particular interest given the apparent recent resurgence in the popularity of proposals to "tax back" insured health care benefits by including them in an individual's taxable income [13,14]. Although such proposals are currently offered as if they were fresh new ideas, in fact they have been around for some time; the Ontario Economic Council itself proposed a detailed scheme for this form of direct charge in 1976 [15].

The Council's proposal involved adding to a patient's income tax liability an amount determined by applying a combination of deductibles and coinsurance to the total health care claims paid on his or her behalf. The deductible level and coinsurance rate were each a function of income and there was a ceiling on the patient's direct financial liability for incurred health care costs which also varied according to income. The income tax return was seen as a vehicle both for the determination of income class and for the payment/collection function.

Our OEC79 analysis examined this specific form of direct charge and two closely related variants. The first was what is currently referred to as the "tax back" approach. Calculation of the health care costs attributed to an individual would be included within the overall assessment of taxable income, without any explicit income class determination. Costs incurred during receipt of care in a given year would be equated with taxable health care benefits and the 'direct' charge for care would then be determined by an individual's marginal tax bracket. A ceiling on taxable health care benefits could be incorporated (similar to the present ceiling on Canada Pension Plan contributions). The main difference between the Council's scheme and this one was that, in the Council's, attributed health care costs led directly to an increase in tax payable (for non-exempt consumers), while in this variant the increase was to taxable income and then only indirectly to tax payable.

The second variant of an income-related direct charge was termed a "health rebate". The idea was that individuals who used no health care during a given period would be eligible to claim rebates when filing their income tax returns; partial rebates would accrue to those who used only limited amounts of service. The rebates would vary by income class, and alternative ways of doing this were identified.

OEC79 [2, p.60-72] provides numerous comparisons of specific and sometimes subtle effects among these three schemes, and between these schemes and the other forms of direct charges discussed above. The important conclusions, however, related to the comparison between this 'family' of direct charge schemes and universal, first-dollar coverage.

In that comparison, the income- and income tax-linked schemes fared not much better than the other forms of direct charges discussed previously. No new incentives for technical efficiency were created, because reimbursement was still uniform across providers. There was no reason to expect significantly larger effects on utilization because the schemes were basically
combinations of coinsurance and deductibles, and the income ceilings and/or accompanying exemptions would still insulate against the deterrence effect of the charges (although supplementary private insurance might be more difficult to integrate with these schemes). And, most important, patients paying the charges would still be exposed to increased financial risk.

Relative to coinsurance, deductibles, and per-service charges, the income- and income-tax linked schemes (especially the Council’s) represented an improvement on the policy objective of wealth transfer. The increased progressivity in the distribution of financial burden enhanced vertical equity (the equitable distribution of burden across income classes. However, the schemes still performed poorly on the criterion of horizontal equity (equal treatment within an income class); within an income class the wealth transfers were still from the sick to the healthy.

Compared to universal, first-dollar coverage financed from general government revenues, however, even the improvement in vertical equity was highly unlikely. It depended critically on the relationship between income and the incidence of illness, and on the progressivity of the revenue sources used to finance the universal, first-dollar coverage plan.9 In general, the income- and income tax-linked schemes envisioned in popular proposals are less progressive than the first-dollar coverage alternative, unless one assumes that health care use is more progressively distributed than tax revenue - which is not the case [16].

Some additional difficulties unique to the income- and income tax-linked schemes (such as the requirement for accurate information on the costs of hospital stays) were noted in the OEC79 analysis, but are not repeated here. The summary in this sub-section should itself suffice to make advocates of this family of direct charges reconsider the alleged merits of the schemes, in particular the misleading and erroneous notion that they improve progressivity.

9 We address this issue elsewhere in the current project [5,6]. *For any level of health care spending, ... user charges shift the burden of costs away from taxpayers generally (both rich and poor) and onto users of services (both rich and poor) ... of course, most people are both taxpayers and patients, so they both gain (as taxpayers) and lose (as patients). Whether they gain or lose overall depends on the amount of taxes they pay and the amount of care they use. In general, wealthy people pay more taxes and sicker people use more care; moreover, wealthy people tend to be healthier, and poorer people sicker. The healthy rich thus stand to gain the most from the introduction of user charges and the sick poor stand to lose the most. Viewed this way, well-intentioned advocates of user charges seem more like the Sheriff of Nottingham than Robin Hood. [5,p.30-31]
2.3.5 Major-Risk Medical Insurance

This form of direct charge was receiving considerable attention in the United States in the 1970s, and was included in the OEC79 analysis partly for that reason, even though it was seldom discussed in the Canadian context. Major-risk medical (MRM) insurance refers to insurance policies with relatively large deductibles (well above the annual health care expenditure of the average individual) which protect individuals only against catastrophic financial losses resulting from the treatment of serious illness. The policies are traditionally sold by private insurance companies who exercise no influence over provider fees. The private company reimburses the patient for all or a portion of expenses above the limit, after these expenses have been paid by the patient. Health care providers are free to charge patients whatever is deemed appropriate.

This means that, unlike all forms of direct charges discussed above, patients may encounter differential charges across providers. This is an important analytic distinction, because the potential for price competition among suppliers is a necessary condition if decisions by patients in response to direct charges are to improve the technical efficiency of the production and delivery of health care. In theory, the choices of informed consumers reacting to price signals (direct charges) allow more efficient suppliers to increase their market shares by lowering their prices.

In practice, however, for the mechanism of consumer-payment\price-competition\consumer-choice\provider efficiency to function, two sets of preconditions are necessary, one concerning market structure and conduct, and the other concerning product characteristics and consumer behaviour.

The first set amounts to the display of price-competitive behaviour. Charges by providers must be truly independent, open, and advertised, without either formal or informal collusion among providers, unlikely though this may be both behaviourally and institutionally [2,p.75]. The second set requires that patients be sufficiently informed to be able to judge the relative merits of the alternative service packages offered at different prices or that regulatory mechanisms exist to ensure adequate quality and safety standards for the services provided.

The OEC79 analysis examined several specific variants of differential (across providers) charge schemes. In the case of MRM insurance and the next two forms discussed below (extra-billing and service repackaging) the charges are provider-determined. In the two final cases below (selective deinsurance and parallel systems) the charges are determined by third-parties other than health care providers.

The assessment of MRM insurance compared to universal, first-dollar coverage found MRM obviously inferior on risk reduction and wealth transfer grounds. The analysis focused on the remaining objectives, however, because advocates of MRM in the United States typically claimed that MRM insurance controlled utilization and expenditures by discouraging unnecessary

---

10 For detailed analysis and references, see OEC79 [2], p. 76-82.
care, and improved efficiency by encouraging patients to price-shop among competing providers.

Analysis of the first claim, regarding utilization and expenditures, found that all of the considerations discussed earlier in the coinsurance context also applied here. For the non-elderly, non-poor, non-hospitalized U.S. population there was some evidence that MRM-type plans deterred some utilization, but the groups of patients in the empirical studies from which these results were derived were far from representative of the average user of health care. Furthermore, there was no evidence to substantiate the specific claim about deterrence being concentrated on only unnecessary services. Finally, it was noted that under MRM-type schemes (unlike under uniform charge schemes), if markets did not function perfectly, providers affected by a deductible-induced fall in demand for their services could offset reduced incomes by raising prices for some or all of the remaining utilization in addition to exercising whatever demand-generating latitude they held.

More general observation on the relative experiences of Canada and the United States post-1971 suggested that the U.S. system, which relied on ‘market forces’ to control costs and improve efficiency, had performed significantly more poorly on these dimensions than the Canadian system, which used centralized provincial bargaining of physician fees and direct provincial control over hospital budgets.

2.3.6 Extra-Billing by Physicians

This form of direct charge allows individual physicians to impose, at their own discretion, extra charges of varying amounts on some patients over and above the reimbursement schedule negotiated with the paying agency (the provincial governments in the case of Canada). Price discrimination of this type allows physicians to charge more to those who are least price-sensitive and less or nothing (additional to the schedule of benefits) to those who are most price-sensitive. In this way, physicians can achieve the maximum possible increase in the average price of their services consistent with little or no reduction in demand for their services.

In the 1970s and early 1980s, extra-billing by physicians who had ‘opted out’ of their provincial medical insurance plans and billed their patients directly was the form of user charge most familiar to Canadians. Subsequent to OEC79, extra-billing became the subject of major public and political discussions [3], which culminated in the 1984 Canada Health Act designed to eliminate the practice.

---

11 For detailed analysis and references, see OEC79 [2], p. 82-90.

12 Patients were then reimbursed by the provincial plans according to benefit schedules previously negotiated between governments and physicians’ medical associations.
The analysis of extra-billing in OEC79 found that it held "little appeal as a tenable instrument of cost control, promotion of efficiency, risk reduction or socially acceptable wealth transfer" [2,p.90]. To the extent that extra-billing physicians were skillful at price discrimination, utilization would remain relatively constant while price, and therefore medical costs (and medical incomes), rose. If physicians had no uniform rate of extra-billing, information on the effective prices for any physician's services were very difficult to collect and comparative price shopping was virtually impossible. Patients were clearly exposed to more financial risk, and costs were shifted from taxpayers to the ill (although if physicians were adept at assessing the income class of patients, it might at least be a progressive transfer).

The OEC analysis proceeded to examine more closely some possible differences between physicians upon which differential reimbursement could be based, but this did not alter the overall assessment, and highlighted other problems which stemmed directly from the structural feature of extra-billing schemes which allowed the very people whose behaviour one might wish to affect with differential prices - individual physicians - to set the prices!

2.3.7 Service Repackaging

This form of direct charge was only briefly developed in the OEC79 analysis. It referred to attempts to "repackage" the health care services in question in order to create a product about which consumers might reasonably be expected to make informed decisions, and to allow differential (provider-determined) charges for this "new" product.

The two examples used in OEC79 were differential dispensing fees in prescription drug insurance plans, and Health Maintenance Organizations (HMOs) in the United States.

In the former, the process of drug dispensing is distinguished from the provision of particular drug ingredients, and the plans pay for ingredient costs in full but only pay a basic minimum toward the dispensing charge. The remainder of the charge is set by each pharmacy and charged to the patient, who is free to choose which pharmacy to use on the basis of dispensing fees as well as other relevant characteristics. In theory at least, there is scope for price competition on the dispensing function among pharmacies to lead to efficiency improvements such as increased use of dispensing assistants. As discussed in OEC79 [2,p.90-91] there are several market structure and conduct conditions that still need to be met in order for potential efficiency improvements to be realized.

In the case of HMOs, two significant changes are happening to the medical services market when HMOs compete with other providers by offering a package of necessary services for a given period (usually a year) at a given price (enrolment fee). First, a new type of health care delivery organization is being created which combines the insurance and service functions. Second, the "product" has been redefined to be a source of care for a particular period rather than specific health services defined, priced, and bought as needed. The new product is one about

---

13 For detailed analysis and references, see OEC79 [2], p. 90-96.
which consumers can inform themselves more easily and accurately. In this context, differential charges between HMOs and fee-for-service-reimbursed (groups of) physicians have the potential to channel patients to the lower-cost providers, again on the assumption that market structure and conduct as well as these product-characteristic/consumer-information conditions are met. (This is, in fact, the cornerstone of the current Clinton health care reform proposals in the United States.)

OEC79 briefly identified the possibility for competition between HMO-style, capitation-reimbursed organizations and fee-for-service practitioners in Canada. The suggestion was made that, "A constructive use of financial incentives might be to pay rebates or to lower premiums to groups of patients who choose service delivery organizations with lower costs" [2,p.93]. But several regulatory obstacles requiring legislative change were also identified.

The overall conclusion on service repackaging was not much different than for other forms of provider-determined differential charges. They increased financial risk for patients and transferred wealth from users of care to taxpayers in general. They might mitigate cost escalation to the extent that the repackaging led to incentives for more effective and efficient care. But it was deemed unlikely that these benefits would materialize without the intervention of some third party to establish differential charges with the appropriate incentives, to prevent collusion among providers, and to monitor quality of care.

2.3.8 Selective Deinsurance

This form of direct charge, and the one which follows (parallel systems) were offered as "potential policy directions warranting energetic further exploration" [2,p.98]. Both envisioned government as the third-party determining the differential charges in the Canadian context.

Selective deinsurance was defined to be refusal to reimburse patients or providers for ineffective services received or rendered. On the premise that it is inappropriate for public health insurance to cover procedures shown to have no clinical effectiveness, such procedures would be "deinsured", thereby permitting providers to level a direct charge on patients equal to 100% of the provider's fee.15

Numerous difficulties with this approach were identified16, and definitive assessments

---

14 For detailed analysis and references, see OEC79 [2], p. 99-104.

15 A more detailed discussion of the grounds for deinsurance appears in one of the papers in our current project [8].

16 The information requirements would be extensive, and included rigorous scientific demonstration that services were in fact ineffective for specific patients in specific circumstances. The main problems of unnecessary
of its impact were precluded by the tentative nature of the policy direction and inadequate information on the range of procedures which might be affected. It did not seem unreasonable to suggest, however, that if the difficulties could be surmounted, the policy would be approximately neutral with respect to both risk reduction and wealth transfer (especially since payment of charges would be unrelated to consumption of effective health care). Effects on aggregate utilization depended largely on assumptions about how providers would react, but the incentives for efficiency were strong and direct.

2.3.9 Parallel Systems

The idea here was that differential charges directing patients to the more efficient alternative might be set by government in situations where a lower-cost, equally-effective delivery alternative existed in parallel with a higher-cost alternative.

If the alternatives were in fact of equal therapeutic effectiveness - an absolutely critical requirement - then the differential charges would in effect be charges for what might be termed consumer preferences for "ambience" (i.e., aspects of the treatment process or setting that were not essential for the therapeutic outcome), something that consumers could reasonably be expected to judge the worth of for themselves.

Several examples were used for illustrative purposes, including immunizations provided by physicians instead of public health nurses, semiprivate/private instead of ward accommodation in hospitals, day- instead of in-patient surgery, public health unit well-baby care instead of care by paediatricians, and school-based children's dental care instead of private dentists. The key servicing were not likely to be in the area of demonstrably useless services, but were more likely to be associated with the provision of sometimes effective services in specific clinical circumstances in which they were not effective, a situation which would call for a much more subtle and operationally complex application of deinsurance. Information would also have to be communicated carefully to consumers to fulfill the product-characteristic/consumer information conditions. Finally, provider co-operation would be critical to the success of the policy. If providers publicly challenged the government's information, or reclassified procedures to avoid 'official' use of deinsured codes while still providing the service, or reacted to lower incomes by generating increased demand for other unaffected services, then deinsurance would be of limited usefulness.

For detailed analysis and references, see OEC79 [2], p. 104-108.
requirement in all cases, however, was that a therapeutically equivalent and accessible public service must exist for which patients would not be charged.\footnote{Recent proposals for user charges for inappropriate use of emergency rooms \cite{17} could also be considered in this category if patients had access without charges (and knew this) to 24-hour service from medical clinics in close proximity to the emergency facility. Analysis of such "steering charges" is not quite so straightforward, however, as we point out elsewhere \cite{8}.} In addition, prospective patients must be aware of this and be convinced of the equal effectiveness of the lower-cost alternative. Therefore, both an active public information campaign and co-operation from providers would again be required.

The potential impacts of this approach were tentatively judged to be very similar to those of selective deinsurance: neutral with respect to risk reduction and wealth transfer (for those using the lower-cost alternatives) and, (in the absence of countervailing shifts in behaviour by providers and with appropriate capacity adjustments by government) positive on aggregate utilization, expenditures and efficiency.

A final variant of the parallel system approach involved incentives for self care. Here the potential policy involved providing positive financial incentives (e.g. stipends) to patients undertaking their own care with supervision at home rather than in a hospital setting, especially in the area of chronic care. Although this approach was not developed or analyzed in detail, its potential effects were considered to be similar to those of the other parallel system variants with the exception of wealth transfer effects, which would depend upon the distribution of income class of patients who received the incentive payments (but might be expected to be positive given the incidence of illness by income class).

2.4 Overview of Policy Impact of Direct Charge Schemes

Figure 2 reproduces the policy impact matrix in OEC79 that summarized our "best guesses" regarding the likely effects of each of the types of direct charges that we analyzed. In general, direct charges appeared to offer few significant social benefits, and frequently appeared to have perverse effects. There seemed to be a natural division in particular between the types of schemes which had at that time been most frequently proposed and implemented (those in Figure 2 down to and including extra-billing) and the more limited range of less familiar, less well-defined alternatives whose operational feasibility remained largely untested.

The conclusions of the OEC79 analysis are best summarized in words from the original monograph:

...the scope for deployment of direct charges as a strategy for cost containment or efficiency improvement is extremely limited...the purported benefits of direct charges to
patients either are absent or seem to accrue primarily to providers... In the present structure of health care delivery, most proposals for 'patient participation in health care financing' reduce to misguided or cynical efforts to tax the ill and/or to drive up the total cost of health care while shifting some of the burden out of government budgets. [2,p.116]

3. Review of Literature, 1979-1993

In this section we briefly review additions since 1979 to the research literature on the subject of user charges and their effects. The review is selective in that it emphasizes empirical studies of charges rather than the commentary or debate about them. The review is limited to Canadian and American studies; however, in the third part of this section we place the North American micro-studies in the macro context of trends in international health care spending and delivery systems among developed countries.

3.1 Canadian Studies

3.1.1 Summary of the Saskatchewan Experience (Beck and Horne, 1980 [21])

Although published in 1980, this article basically summarizes and synthesizes several earlier analyses of the effects of user charges in Saskatchewan from 1968-1971 that were published by the same authors during the 1970s and were incorporated into the OEC79 analysis. However, the article does include some previously unpublished results concerning the post-copayment period and concerning diagnostic- and procedure-specific analysis based upon the individual rather than the family.

In Canada, for example, commentaries about one specific type of charge—extra billing by physicians—proliferated until the mid-1980s, fuelled first by a review of Medicare by Justice Emmett Hall [18] and second by the introduction and passage of the Canada Health Act [19] in 1984 (and provincial actions to comply with the Act). References for many of these commentaries are included in the Bibliography for the current series of papers [20]. The number of published studies of the effects of extra-billing was, however, quite small. We review the major studies of which we are aware.

Literature concerning the deployment and effects of user charges in developing rather than developed countries is increasing rapidly. The quite different context of the developing world severely limits the applicability of empirical studies for present purposes, however, and they are not discussed here.
The main new finding was that the removal of the user charge for hospital inpatients ($2.50 per day for the first thirty days and $1.50 per day thereafter to a maximum of ninety consecutive days) did not increase utilization, nor was there a length of stay effect (increase) resulting from the removal of charges, even on a diagnosis-specific basis. This confirmed the insensitivity of hospital utilization to user charges in the authors' previous work.

The multivariate analysis of the utilization of physician services before, during, and after copayment confirmed an average copayment effect (decline) for the three-year period of 7.7%; however, as had been reported by the authors previously, this effect was the result of a much larger decline in utilization (approximately 18%) among lower-income groups, accompanied by some increases in utilization among higher-income groups. The authors concluded that the copayment effect for physician services "undoubtedly" contained a supply response by physicians to decreased demand by patients and that therefore the observed aggregate copayment effect on utilization "likely understates the behavioural response of consumers to direct charges" [21,p.797].

3.1.2 Extra-Billing In Ontario (Stoddart and Woodward, 1980 [22])

Commissioned by Health Services Review '79 (the second Hall Report [18]), this randomized telephone survey of 1,769 households investigated the impact of opting-out²² and

---

²¹ In fact, the probability of at least one family member's admission (standardized for family characteristics) fell after the removal of the charges.

²² An earlier study by Wolfson and Tuohy [23], published in 1980 but which drew on data from 1976-1979, examined the practice behaviour of opted-out versus opted-in physicians in Ontario; however, the focus of the study was not on extra-billing per se, nor on the impact of opting-out and extra-billing on utilization. Consequently, it is not reviewed here. A main finding of this study was that "once other differences between opted-in and opted-out physicians were taken into account there is no indication that opting out in itself made a difference to their practice behaviour" [23,p.viii]. They also found that 11.5% of contacts with opted-in physicians were seen by these physicians as unnecessary compared with 8.5% of patient contacts for opted-out physicians; however the definition of "unnecessary" was left to the individual respondents, which rendered interpretation of these data difficult. (Elsewhere [5] we discuss in detail the difficulty of defining patient-initiated unnecessary use.) Wolfson and Tuohy did, however, express serious concern about the effects of opting-out and extra-billing, stating "The major problem posed by opting-out now is that it erects financial
extra-billing by Ontario physicians on patients’ access to care and attitudes toward the Ontario health system. The survey was conducted in early 1980 in four Ontario counties which were among those with the highest percentages of opted-out general practitioners in Ontario at the time. It included respondent groups with and without experience with an opted-out physician who extra-billed.

The study found that extra-billing did have a significant effect on access to care, especially for lower-income households. Findings included the following:

1) 19% of respondents whose physician extra-billed reported that they had gone to their physician less often in response to extra-billing.
2) 14% of respondents whose physicians extra-billed indicated that there had been at least one occasion on which they felt that they or an adult member of their family should have seen a doctor but did not,
3) the poor were significantly more likely than the non-poor to report having reduced their utilization and/or delayed seeking care in response to extra-billing, and
4) in the absence of extra-billing no difference was found in the proportion of the poor and non-poor who reported delaying or failing to see a doctor because of cost.

There were two important caveats to the study. First, because actual utilization data from the Ontario Health Insurance Plan were unavailable to the investigators, it was not possible to confirm self-reports of decreased utilization in response to extra-billing. Second, no information was available (other than respondents' views) on the necessity or urgency of the utilization which was reported to have been foregone or delayed. Nevertheless, the study results clearly indicated that respondents perceived extra-billing to be a significant financial burden and a barrier to access, and one that had a significantly greater impact on lower-income households.

3.1.3 Hospital User Fees in New Brunswick (Duffey, 1983 [24])

In an article arguing that better "marketing" is needed on the part of governments and hospitals in order to make user fees acceptable to the public and less difficult to administer, the author reports some limited information on the effects of hospital user fees in New Brunswick during 1979-1980.

User fees consisted of a $10 admission charge ($4 for those over 65 years of age) for inpatient stays and a $6 fee per visit ($2 for those over 65) for emergency room and outpatient services. Exemptions were "quite numerous and there were variations in the way individual barriers to care." [23,p.194]
hospitals interpreted them and this lead (sic) to administrative problems" [24,p.30]. Rebates were available for individuals (families) who had paid more than $150 ($225) during a seven-month period.

The author describes the experience of the Moncton Hospital and the provincial political context. The only empirical finding presented is that the user charge policy, which was in practice apparently applied differently by different hospitals, resulted in a reduction of "demand for outpatient services" of 12.7% in the first seven months after user fees were imposed, compared to the same seven months of the previous year.

As the author acknowledges, many questions remain unanswered about the ten-month experience with hospital user charges. It is not known, for example, whether patients increased their use of physicians (there were no user charges for physician office visits); the effect on total utilization of services is therefore unclear. It is also not known whether the user charges affected the use of necessary services (although the author’s opinion is that they did not).

The user fees were removed in 1980, in part because the financial position of the province had improved considerably, but also because of "their limited revenue impact, inequalities due to income differences and administrative complexity" [24,p.27]. The author also points out that there was widespread opposition to the user charge policy in the provincial legislative, from unions (including the nurses’ union), and from other public groups. She conjectures that a political imperative was at work: "Perhaps the party in power would not have wanted to face an election with user fees still in place" [24,p.33].

3.1.4 Billing Above Tariff In Nova Scotia (Brown and Hicks, 1984 [25])

As part of a more general, descriptive study on billing above tariff (extra-billing)23 by Nova Scotia physicians from 1969/70 - 1982/83, the investigators estimated the effects of billing above tariff on the utilization of insured physician services within cohorts of poor and nonpoor persons during the early years of Medicare. What the authors refer to as "implied" price elasticities of utilization are reported for physician services utilization by poor and nonpoor Nova Scotians (excluding infants and the elderly) for the first year of Medicare, 1969/70, when billing above tariff was at its highest level.

The investigators found that billings above tariff had "a measurable, but small, impact on

---

23 Nova Scotia physicians were not required to opt-out of the provincial Medicare plan in order to bill patients for amounts in excess of the provincial reimbursement rates for their services. They submitted claims cards to and were paid by the provincial plan. They were permitted to bill patients for additional amounts provided they notified the patient in advance, the patient consented to the extra-billing, and the extra amount billed was shown on the claim card.
utilization of insured physician services, after standardizing for codeterminants of utilization such as age, sex, location, access to physicians, comprehensive first-dollar medical care insurance coverage and family income" [25,p.259]. Findings included the following:

1) for persons with some utilization, billings above tariff charges (which averaged less than 6% of the value of insured services utilized) were estimated to have resulted in a reduction in utilization levels of less than 1% for both poor and nonpoor populations,

2) price elasticities were consistently higher for the poor than the nonpoor, although they were very inelastic for both groups, and

3) absolute reductions in utilization of physician services were less for the poor than the nonpoor, but only because the mean level of billing above tariff was markedly lower for the poor.

The authors acknowledged that there were a variety of factors affecting the estimated elasticities which prevented their interpretation as standard price elasticities. The authors also identified several factors to account for the difference between the small utilization effect of extra-billing in their study and the larger utilization effect of deterrent fees in Saskatchewan. Most important was the fact that only a very small minority of Nova Scotia utilizers experienced substantial direct charges comparable in size to those in Saskatchewan.

3.1.5 Extra-Billing in Alberta (Plain, 1984 [26])

This study provided detailed statistical profiles of the extra-billing activity of Alberta physicians during the period 1979-1982 and addressed "certain of the socio-economic effects stemming from extra-billing" [26,p.26]. On the latter subject, data released by the Alberta government showed that from 1979-1981, lower-income and elderly Albertans were extra-billed by their physicians. This contradicted claims by the medical profession at the time that physicians engaged in selective extra-billing which exempted the poor and seniors. The author also pointed out that such extra-billing practices violated the social norms of the province regarding the groups deserving of exemption status, as indicated by provincial policies regarding exemption from (or subsidy of) payment of health insurance premiums.

"For example, annual billings above tariff for females within the Nova Scotia nonpoor cohort averaged $3.00; the charge for a single office visit in Saskatchewan was $2.50. As a percent of the value of services utilized, billings above tariff in Nova Scotia were 5.7% for the nonpoor and 2.2% for the poor in 1969/70; in Saskatchewan user charges for insured physician services were about 33%." [25,p.267]
Data from 1979-1981 on the percentage of registrations extra-billed by type of registration showed that 11-15% of those registrants receiving complete subsidy of premiums were extra-billed by their physicians. For social allowance recipients, partial subsidy recipients, and those 65 and over the percentages were 4-6%, 10-14%, and 3-5%, respectively. (The figure for non-subsidized registrants was 18-24%.)

The available data did not permit the author "to carry out the econometric analyses needed to measure the degree to which extra-billing has reduced access to and utilization of medical services of various classes of consumers" [26,p.57].

3.2 Studies from the United States

3.2.1 The Rand Health Insurance Experiment (Manning et al., 1987 [28], Brook et al., 1983 [29])

The American literature on user charges continues to be considerably larger than its Canadian counterpart, which is not surprising since the U.S. health care system relies much more extensively (and unsuccessfully - see section 3.3 below) on user charges of various types to control expenditures. We wish to thank Thomas Rice for assistance with the identification and digestion of the American literature. Rice [27] has recently completed a comprehensive review of the U.S. literature on the impact of cost-sharing (copayment) for several population groups including the non-elderly, the elderly, Medicaid recipients, uninsured persons, and enrollees of Health Maintenance Organizations. Rice discusses the methodologic problems which have weakened many of the studies, particularly those which have used cross-sectional data. The studies included in our review of the U.S. literature are those since 1979 which Rice concludes have the best internal validity (i.e. findings are valid for the setting and population studied). However, as Rice notes, and we discuss further below, in several cases important questions remain about the external validity of the studies (i.e. the generalizability of their findings to other settings, populations or health care systems).

The Rand Health Insurance Experiment generated numerous research publications by many different investigators on a wide range of outcomes including utilization, health care expenditures and health status. Manning et al. [28] and Brook et al. [29] are perhaps the two most significant and widely cited articles. See Manning et al. for a detailed description of the experiment.
The Rand Health Insurance Experiment (HIE) is the largest (and at $136 million in 1984 U.S. dollars the most expensive) study of user charges ever conducted. A true experiment, it randomly assigned 5,800 individuals in six U.S. sites for a three- or five-year period to one of 14 health insurance plans involving fee-for-service physicians and different combinations of coinsurance and maximum limits on out-of-pocket expenditures. Coinsurance rates of 0% ("free care"), 25%, 50%, and 95% were used, and out-of-pocket maximums were either 5%, 10% or 15% of family income, up to an annual maximum of $1,000, beyond which the insurance plan reimbursed all covered expenses in full. The plans provided coverage for a comprehensive set of benefits including physician services, hospital services and prescription drugs. Significant exclusions from the study population were the elderly (62 years of age or older at the time of enrolment), the wealthy (the 3% of the study population with 1973 income in excess of $25,000), and some disabled (those eligible for the Medicare disability program).

A full review of all results on all dimensions of the HIE would be a paper in itself, and a lengthy one at that. Here we highlight the HIE findings in the areas with the most relevance to the current Canadian debate. These concern the impact of cost-sharing on overall use of services and expenditures, the use of necessary versus unnecessary services, health status, and the extent to which cost-sharing disproportionately affected lower-income groups.

Findings of the HIE included the following:

1) Although differences in the maximum limits for out-of-pocket expenditures did not appear to affect utilization, coinsurance rates did have a significant effect. Per capita annual expenditures in the free care group were 23% higher than those for the 25% coinsurance group, and 46% higher than those for the 95% coinsurance group [28].

2) Price elasticities of demand were approximately -0.2, with generally similar

Additional studies are cited in the text of this subsection, and further citations are available in Manning et al. and Brook et al.

27 One plan provided free inpatient care but had a $150 deductible per person for outpatient services in order to investigate the effect that copayment had on the substitution of inpatient for outpatient care.

28 An "elasticity" measures the responsiveness of one variable to changes in another variable. In this case, for example, a price elasticity of -0.2 indicates that a 10% increase in price leads to a 2% decrease in the
impact for both inpatient and outpatient care (except in the case of inpatient care for children, where cost-sharing did not deter inpatient paediatric stays) [28],

3) Utilization reductions were almost entirely the result of individuals reducing the number of episodes of care-seeking, rather than the result of decreased utilization per episode of care [30],

4) Free inpatient care was not substituted for cost-shared outpatient care [28],

5) Coinsurance reduced the demand for care in situations where care was likely to be highly effective as much as it did in situations where care was deemed to be only rarely effective [31]. Coinsurance was therefore not selective in reducing the use of only less necessary care. Nor did it selectively reduce inappropriate hospitalization [32],

6) Adults in the free care group experienced better health outcomes than those in the coinsurance groups in three areas - diastolic blood pressure [29], corrected vision [33], and the risk of dying for those at elevated risk [29]; however, their experience was not better on several self-assessed health measures including physical functioning, role function, social contacts, mental health or health perceptions, and free care did not appear to reduce weight or cholesterol levels [29,34]. Individuals in the free care group were more likely to reduce smoking, maintain a low-salt diet and comply better with medication regimens [35],

7) Among low-income families, children in the coinsurance group used only 56% as much care deemed to be highly effective as their counterparts in the free care group. (For adults the figure was 59%) [31], and

8) Low-income individuals at elevated risk benefited most from free care, and low-income people in poor health who received free care experienced the largest reduction in serious symptoms [29,36].

This list of results is only a fraction of the total results reported from the HIE. Readers interested in a more complete understanding of the experiment are directed to the studies referenced above and additional references therein.

Even with such an expensive, complicated, and time-consuming experiment, there remain significant methodologic difficulties which affect the interpretation of the results on important questions. For example, as Rice [27] points out, even though coinsurance reduced utilization and expenditure for all income groups, it is difficult to assess the differential impact of alternative quantity of care demanded. In other words, the demand for care is relatively insensitive (or inelastic) to changes in the price of care.
coinsurance rates by income level because the design of the experiment set a limit on out-of-pocket expenditures for lower-income people that was lower than the limit for upper-income people. More significant are the difficulties associated with interpretation of the results on health outcomes over a relatively short period and in the presence of receipt of both necessary/appropriate and unnecessary/inappropriate care by patients in most groups.

The most significant limitation of the HIE, however, was embodied in the basic structure of the experiment. Although the HIE produced the most rigorous results to date on how individuals and families react to cost-sharing, it was from the outset incapable of estimating the effect of cost-sharing on an entire population because it could not assess how the health-care system would react. In other words, what would be the effect on utilization, expenditures and health if all citizen-patients faced coinsurance, deductibles or some other form of user charge?

Only a very small percentage of the population in any one site participated in the HIE and therefore most physicians had a negligible number of (high coinsurance) HIE patients in their practice. Similarly, individual hospitals had only a small proportion of their overall utilization affected by the experiment. Therefore the effects of cost-sharing on health care providers' incomes and revenues were insignificant.

If, on the contrary, all physician and hospital utilization by all patients in the sites had been subject to cost-sharing, and all individuals responded as the HIE coinsurance groups did, then health care providers would have experienced a 15% - 30% decrease in (gross) incomes/revenues. This would have almost certainly elicited some supplier response, but how large, and of what type? These are critical questions for policy-makers considering the introduction of user charges. Unfortunately the HIE cannot provide answers.

3.2.2 United Mine Workers Health Plan (Fahs, 1992 [37]; Roddy et al., 1986 [38])

After 25 years of first-dollar coverage by their health insurance plan, members of the United Mine Workers of America Health and Retirement Fund were subjected to copayment in 1977. For a brief, five-month period, the copayment consisted of a $250 annual inpatient hospital deductible and 40% coinsurance on physician and most hospital outpatient services up to a maximum of $500 per family. These were subsequently reduced to a flat user fee of $7.50 per physician visit, with no cost-sharing required for hospital care other than physician visits.

An early analysis by Scheffler [39] of the immediate pre- and post-copayment utilization by non-elderly members showed dramatic decreases in utilization of both inpatient and outpatient services, but as the author noted, several other factors in addition to copayment may have affected the results. A later analysis by Roddy et al. [38] examined utilization by retired mine-workers of pre-Medicare-eligible age and their dependents for two years following the introduction of cost-sharing, confirming Scheffler's analysis and finding results for the first year that were generally consistent with the HIE results. For the second year of copayment, however, the investigators found that use of ambulatory care increased and returned to pre-copay levels, leading them to suggest that copayment effects for ambulatory care may be short-lived. More detailed analysis indicated that the increases were greater for men and for beneficiaries 45-64
years of age, and for visits of a less discretionary nature (visits that were neither for acute, self-limiting conditions nor for preventive care).

Fahs' [37] analysis is significant because it examines the effect of cost-sharing by patients on physician behaviour (one of the 'holes' in the Rand HIE). She examined the utilization by all patients in a large, multispecialty group practice in western Pennsylvania, for the year before and two years after the introduction of cost-sharing for mineworkers. Over 80% of the patient population of the group practice was made up of two groups: mineworkers and their families, for whom cost-sharing was introduced, and steelworkers and their families, whose health insurance benefits did not change during the 1976-1979 study period.

Fahs found that "when the economic effects of cost sharing on physician service use are analyzed for all patients within a physician practice, the findings are remarkably different from those of an analysis limited to those patients directly affected by cost sharing" [37,p.26]. The physicians reacted to the introduction of cost-sharing for their mineworker patients by increasing inpatient lengths of stay and prices for ambulatory care for their non-mineworker patients. In addition, the physicians began recommending longer intervals between follow-up visits for mineworker patients, but shorter intervals for other patients. Overall, expenditures per episode of inpatient care rose by 19% and total expenditures by 7% for non-mineworker patients. Fahs concluded:

Thus, the results suggest that increasing cost sharing among large groups of patients may be less effective as a tool to reduce total health care expenditures than has been implied by studies that omit the effect of cost sharing on physician practice patterns. It appears from this analysis that compensatory actions will be taken by physicians following the reduction in benefits by a large insurance carrier. [37,p.39]

There are, as usual, questions about the generalizability of this study (of a single group practice with a specific patient population of miners and steelworkers in a particular geographic region), and the author acknowledges some of them. Nevertheless, it provides an important caution to analysts and policy-makers working with results from studies of user charges in which only the responses of patients - and not physicians - have been considered.

3.2.3 Other U.S. Studies

Rice [27] judges the Rand HIE and the United Mine Workers experiences to be two of the three sets of studies which provide most of researchers' current understanding of the impact of cost-sharing, especially among the non-elderly. The third set of studies analyzed the introduction of coinsurance to a prepaid medical plan for the faculty and staff of Stanford University in the late 1960s. This set [40-42] was incorporated into the OEC79 analysis.

In his review, Rice also discusses studies of the impact of cost-sharing on the elderly [43-47], Medicaid patients [48-52], and HMO enrollees [53-55]. Studies on the elderly consistently found that cost-sharing substantially reduced utilization and costs; however, the studies were
subject to major methodologic limitations, including selection bias. Studies of the early 1970s California experiment with limited cost-sharing for Medicaid patients (poor, disabled, and blind individuals meeting state eligibility standards) were incorporated into the OEC79 analysis [48-50]. The other two studies on Medicaid patients, both considered by Rice to be of minor significance, showed that copayment resulted in some decrease in utilization of contraceptive services [51] and well-baby care [52].

Another set of studies, by Lurie et al. [56,57], investigated the effects on California Medicaid patients of the loss of health insurance benefits in 1982 when the state terminated its Medi-Cal program and transferred responsibility for care of 270,000 medically indigent adults to the county system, where patients typically faced user fees of $20-$30 per outpatient or emergency department visit. Follow-up surveys at six months and one year found a significant deterioration in the health status of a group who had to pay for their care compared to a similar group who did not have to pay.

A final set of studies by Cherkin et al. [53-55] recently examined the effects of copayment in the Group Health Cooperative of Puget Sound Health Maintenance Organization. In 1984 state employees became subject to a $5 copayment for ambulatory care visits while federal employees (who were covered by a different contract) did not. The copayment resulted in decreases in utilization in the first year of 11% for primary care visits, 14% for physical examinations, 3% for specialty care visits and 20% for primary care visits by patients taking cardiovascular medications. Although there were marked differences in the impact of copayment on subsets of the enrolled population (e.g. women under 40 reduced primary care visits twice as much as men), the investigators found no relationship between income and responsiveness to copayment. They considered this to be "not surprising" because "few persons in the study sample could be considered to have very low incomes" [55,p.40].

---

29 All studies investigated the effect of cost-sharing by examining the impact of having more complete insurance coverage on utilization. If those with more complete coverage use more services, it could be due to the lower cost-sharing levels. But it could also be because sicker people are more likely to purchase the supplementary, more complete insurance.

30 They too were affected by methodologic limitations of the study designs; however, the results suggested that low levels of cost-sharing for outpatient physician services reduced utilization of those services (especially for some preventive services) and were associated with increases in hospital utilization.
3.3 The International Context

In OEC79 we presented a comparison of total health expenditure as a percentage of gross national product (GNP) for Canada and the United States from 1960-1976 [2,p.6]. In 1960 Canada's percentage (5.6) slightly exceeded that for the U.S. (5.3); however, by 1976 the U.S. percentage (8.6) exceeded that for Canada (7.1) by a wide (and growing) margin. During the intervening years, the most significant development in comparative health policy was the completion by Canada in 1971 of a universal, publicly-financed and administered system of first-dollar medical and hospital insurance coverage using a 'single-payer' (provincial government) model. In contrast, the United States had continued to use a non-system of multiple-payer, private insurance coverage supplemented by public programs for the aged (Medicare) and the poor (Medicaid). The private schemes relied extensively on coinsurance, deductibles and other forms of user charges to control utilization and expenditures.

The data vividly illustrated two points: i) that the absence of user charges (i.e. the Canadian approach) did not lead to uncontrolled expenditures (Canada's percentage actually stabilized after 1971), and ii) that the presence of user charges (i.e. the American approach) did not guarantee expenditure control.

With another decade of data on international trends in health care delivery and spending now available, it is possible to re-examine the U.S. - Canada percentages and to place them in the broader context of other industrialized countries' experience. In both the United States and Canada health care spending has increased, and the gap between them has continued to widen. In 1991, the most recent year for which complete (though provisional) data are available [58], national health expenditures represented 9.9% of gross domestic product (GDP) for Canada and 13.2% for the United States.

Elsewhere, Evans [59] has examined the growth in Canadian health care spending in detail, and in an international context. He points out that, although 1) Canadian efforts at expenditure control have been successful when compared with the U.S.; 2) costs are not 'out of control' in Canada (unlike the U.S.); and 3) public health insurance in Canada has been associated with slower rather than faster rates of expenditure growth, Canadians should not be either smug or complacent. The Canadian system has been unable to keep rates of growth of health care spending in line with the much slower rates of economic growth generally throughout the 1980s. This represents a serious political as well as economic problem.

Other countries have managed better performance. As Table 1 based on data for the 24 industrialized member countries of the Organization for Economic Cooperation and Development (OECD) [60] shows, most countries stabilized their health expenditures as a share of GDP before or during the 1980s. Although most countries have some form of patient cost-sharing, at least for some types of services, as a component of their financing mechanisms, it is fair to say that none (outside of the U.S.) relies on cost-sharing as an important or critical part of an overall
expenditure control strategy. The possible exception is France, one of the least successful European countries at controlling expenditures during the 1980s.

Furthermore, although health care reform has been high on the agenda of most European countries recently, interestingly it has come largely after the stabilization of expenditures [62]. And, although "... governments resorted to some extent to increased cost sharing, ... most of the reforms aimed at containing costs were concentrated on the supply side." [61,p.8]. The reforms were primarily directed toward creating supply-side incentives or authority to improve efficiency and/or to increase the responsiveness of suppliers to consumers [61-64]. In effect, most countries have found ways of tightening controls on the behaviour of suppliers, whereas cost-sharing loosens such control.

It should come as no surprise that the focus of health care reform internationally -- and even in the United States with President Clinton's proposals -- is now firmly on the supply-side. The health services research literature of the past three decades has, for most analysts and policymakers, convincingly demonstrated that patterns of utilization of health care respond to incentives addressed to providers, and that it is the attitudes, knowledge and behaviour of providers, rather than consumers, that are the key determinants of utilization. Consumers may initiate utilization,

31 The minor role played by user charges outside the United States -- and not the presence or absence per se of the charges -- is the important point. For example, it is wrong to say that Canada does not have user charges while everyone else does. Canadians face a variety of user charges for health care services such as drugs, ambulance services and other services beyond medical and acute hospital care. (We discuss their rationales, or lack of such, elsewhere [8].) And not all other OECD countries have charges for hospital and medical care, although they may have them for other types of services. It is not our intent to provide an international survey of charges, but rather to suggest that statements about the existence of user charges in countries' health care systems need to be service-specific in order to be meaningful.

32 An analysis for the OECD by Hurst points out that France permitted certain doctors to elect for extra-billing at the beginning of the decade, but by the end of the decade it had become difficult for patients in large cities to find specialists who did not extra-bill. Viewing this as a threat to access by low-income groups, the government in 1990 placed limitations on future elections for extra-billing by physicians. [61,p.139].
but providers largely control it. 33

4. Conclusion
In section 2.1 above we remarked on the striking similarity of key elements of the health policy context in 1979 to those of today. And, indeed, the adage "The more things change, the more they stay the same" applies also to the conclusions of our current examination of the effects of user charges on the five policy objectives (risk reduction, wealth transfer, levels and patterns of utilization, efficiency) in the OEC79 policy impact matrix presented earlier in Figure 2.

The literature and experience of the past fourteen years on balance does not change the assessments offered in that matrix; in fact, it reinforces many of them and confirms several of the predictions and analytic speculations contained in the OEC79 monograph [2].

For the largest group of specific types of charges in Figure 2 -- including coinsurance, deductibles, per-service charges, income- and income tax-linked charges, major-risk medical, and extra-billing -- there is still no reason to expect any positive contribution to the achievement of any of the policy objectives in the matrix. 34 On the contrary, over time it has become even clearer that:

1) charges do not lead to selective reductions in utilization of only unnecessary or less necessary services, but affect the utilization of needed services as well (while much ineffective care continues to be delivered),

2) the distributional effects -- both financial and health effects -- of charges are potentially quite serious for certain groups in the population, especially as health care use becomes more concentrated within smaller and very ill segments of the population, 35

3) analyses of the effects of user charges must take into account the significant,

33 Recall that in the Rand Health Insurance Experiment, for example, almost the entire effect of cost-sharing was accounted for by a reduction in the number of episodes. Once the utilization process had been initiated, however, utilization per episode changed little in response to cost-sharing [29].

34 What has changed, at least in Canada it seems, is the importance that government policy-makers attach to the objective of generating revenue, regardless of the effects on the other objectives. This is one of the very few objectives that user charges will achieve, as we point out elsewhere [4,5,7].

35 For further discussion on this, see Evans, Barer and Stoddart [16].
offsetting response of suppliers to reductions in patient-initiated utilization, and

4) charges are neither a necessary nor a sufficient condition for overall cost control in health care systems.

The second group of types of charges in the lower part of the policy impact matrix -- including service repackaging, selective deinsurance, parallel systems and incentives for self care -- in 1979 represented largely untested alternatives with the potential, in theory at least, to contribute positively on several policy objectives. Our caution in OEC79 was that the preconditions for the success of these schemes (identified in sections 2.3.7 - 2.3.9 above) would likely be very difficult to establish. This has in fact been the case and, although some recent reforms in both North America and Europe use some of these forms of charges to varying extents, it seems fair to say that their success is still -- as it always has been -- in the future.36

Where, then, does this review leave the current Canadian debate about user charges? The short answer37 is "where it always has -- or should have -- been", which is to say focused on the distributional effects of charges, and the social values which determine an individual's or a society's normative judgement of the acceptability of such effects. In our view, the experience and learning since 1979, about both user charges and the determinants of health care utilization more generally, make it clearer than ever that the user charge debate is about cost-shifting rather than cost control, and about the generation of revenue rather than the improvement of effectiveness and efficiency.

Our 1979 conclusion, noted earlier, is worth repeating: "In the present structure of health care delivery, most proposals for 'patient participation in health care financing' reduce to misguided or cynical efforts to tax the ill and/or to drive up the total cost of health care while shifting some of the burden out of government budgets." [2,p.116]. Another look at the literature, fourteen years later, has not changed that conclusion.

---

36 Canadian policy-makers have seriously engaged only in the selective deinsurance variant, and even here the principles underlying proposals for deinsurance have typically not been made as explicitly as they might be [8]. In particular, policy-makers have thus far avoided the so-called "competitive approach", while monitoring its disappointing lack of success to date in the United States.

37 The entire collection of papers from the current project [3-8, and this paper] is the long answer, although the main themes running throughout are the same.
References


