

CHAPTER 12

PUBLIC INVESTMENT PROGRAMS IN PREVENTION AND RESEARCH

INVESTMENT ACTIVITIES WITH DIVERSE AND NON-MARKETED BENEFITS

Many of the general issues outlined in the previous chapter emerge in specific, and particularly intractable, form in the fields of illness prevention and health care research. Both represent investment activities, in that resources are given up in the present, in anticipation of future benefits in the form of either reduced incidence of illness or injury, or enhanced capability of responding to them. And a special public interest is recognized in each of these fields. Public health activities and public subsidy to, as well as actual conduct of, health care research, predate public intervention in the financing and delivery of health care generally, and remain largely independent programs.

Canadian public health and research activities in 1982 accounted for \$952.8 million and \$327.1 million respectively, or about 3.2 percent and 1.1 percent of total health spending (Canada, Health and Welfare Canada n.d. [1984]). Public health activities, however, are not restricted to prevention, nor is prevention solely the province of public health. Indeed it is somewhat difficult to establish a precise definition of preventive activity, since both curative and to a lesser extent "carative" services may not only alleviate a present situation but also prevent it from deteriorating. Research activity is also included in other sectors of the health economy. A component of prescription drug prices covers the costs of research by private firms, and research and development expenditure by private equipment manufacturers similarly enters the prices of their products, which in turn show up in expenditures by hospitals, private practices, or other providers who purchase the equipment. And of course research outside the health care sector itself, in universities, government, or private industry may have significant effects, intentional or otherwise, on the health care system.

Going beyond Canadian health care expenditures, research in other countries generates new knowledge available in Canada, external effects across national boundaries, such that the link between resources *devoted* to research in Canada and new knowledge *available* for use in Canada is rather tenuous.¹

The arguments for public intervention in the funding and direction of prevention and research follow the same lines as the general arguments for public intervention in health care developed above, though they have some distinctive features. First and foremost, as illustrated by the permeability of national boundaries in allocating the output (as opposed to resource input) of research activities, is the significance of external effects. Costs incurred by particular individuals or groups for both prevention and research generate benefits some or most of which accrue to other members of society, within or beyond national boundaries. Private decision-makers will (quite rationally) fail to take account of these external benefits-, they will forego those investments for which (in terms of equation 11-1) costs exceed benefits when summed over a single value or subset of k , but benefits would exceed costs if all k were counted. Accordingly, optimal levels of activity require that the state either subsidize the private decision-maker -- lower costs to the subset of k thus represented -- or carry out the privately unprofitable activity.

As noted in chapter 3, this is the most ancient of the justifications for public intervention in health care, arising particularly in the context of contagious diseases. It underlies not only public immunization programs, which amount to lowering the cost to a private individual of activities with more general public benefits, but also compulsory regulation of private actions, paternalistic or otherwise.²

THE "NEW PERSPECTIVE": A PUBLIC INTEREST IN PRIVATE LIFESTYLES?

Beyond the field of communicable disease, however, the externality justification for preventive programs becomes much more questionable. My immunization status may influence your risk of infection, but my hardened arteries do not affect your risk of heart attack. Yet the focus of preventive activity has largely shifted, within the last decade, from communicable disease to problems of the hardened artery type. The *New Perspective on the Health of Canadians* (Lalonde 1974) assembled extensive data on mortality and morbidity patterns, showing that the most serious threats to health are now illnesses associated with particular unhealthy lifestyles and environments, against which the effectiveness of clinical intervention is distinctly limited (and very expensive). This document was part of a general shift in attitude, a great resurgence of interest in preventive activity in the broadest sense, as likely to be of greater effect than attempts at after-the-fact care in dealing with heart disease, accidents, or cancer. But this shift also undercut the traditional justification for public intervention in preventive activity. If my lifestyle affects my health, what concern is that of yours? Indeed, one can construct fairly elaborate formal economic models of the consumption of preventive services as a purely private activity. Such models are distinctly implausible, and seem to miss most, if not all, of the interesting policy issues surrounding prevention. Nevertheless, they raise an important question, to which there are several different answers: Why is prevention a public issue at all?

One set of answers involves a reinterpretation of the external effects involved. In motor vehicle accidents, as obvious examples, one person's behaviour significantly affects the health of others. Public programs to encourage safe driving, by exhortation, regulation, or punishment, thus fit comfortably within the externality framework. The linkage is directly from A's behaviour to B's health, not via A's health as in the communicable disease case, but it is no less strong for that.³

Environmental hazards to health similarly fit into the external effects category, insofar as the activities of one person, group, firm, or industry generate threats to the health of others. The role of the state in controlling environmental hazards is a natural extension of its responsibility in the field of communicable disease, except that in the environmental case the polluter is usually a net beneficiary from the process of hazard creation. Thus prevention of illness arising from unhealthy environments will typically involve public regulations or clean-up projects which may be globally beneficial but will impose net losses on some subset of the population. This may raise difficult issues of compensation. It also generates political conflict at a number of levels -- including within the evaluation process itself -- which will delay or deflect preventive activity.

A further twist arises in the area of occupational health and safety -- the problem of defining the boundaries of "the environment." If one conceives of workplaces as settings which workers enter as part of a fully informed and voluntary decision, explicitly accepting all the positive and negative consequences therefrom, then the health consequences, risk of present accident or future illness or disability, are merely part of the employment contract. Public intervention, by regulation, requires some justification in terms of incompleteness in this contract. Systematic

biases in information available to workers, or imperfect competition in labour markets resulting in a restricted range of characteristics of the employment setting offered, may undercut the assumed voluntary and fully informed nature of occupational choice.

Another form of externality is pecuniary, not physical; A's behaviour imposes monetary costs on B insofar as B must share the costs of health care generated by A's illness. For this purpose it is immaterial whether A and B participate in a public health service, a public health insurance program, or a private health insurance program with or without public subsidy, so long as A's behaviour affects B's costs. Collective Bs then have a justification for trying to influence A's behaviour.⁴

Beyond the pecuniary externalities is the more general social interaction, altruistic or paternalistic, which rests on the assumption that members of society care about each other, and about each other's health. Highway safety programs are not indifferent to one-car, one-occupant accidents, and suicide is regarded as something more than an individual consumption (investment?) decision. This argument is particularly relevant to preventive programs directed at children or other persons in a dependent condition; there is a collective responsibility to oversee the decisions made by particular others on their behalf. But it applies between independent (?) adults as well; if A cares enough about B, or B's health, to subsidize health care consumption, then presumably her interest is equally legitimately reflected in public programs to influence B's behaviour or otherwise reduce the probability of B's falling ill.

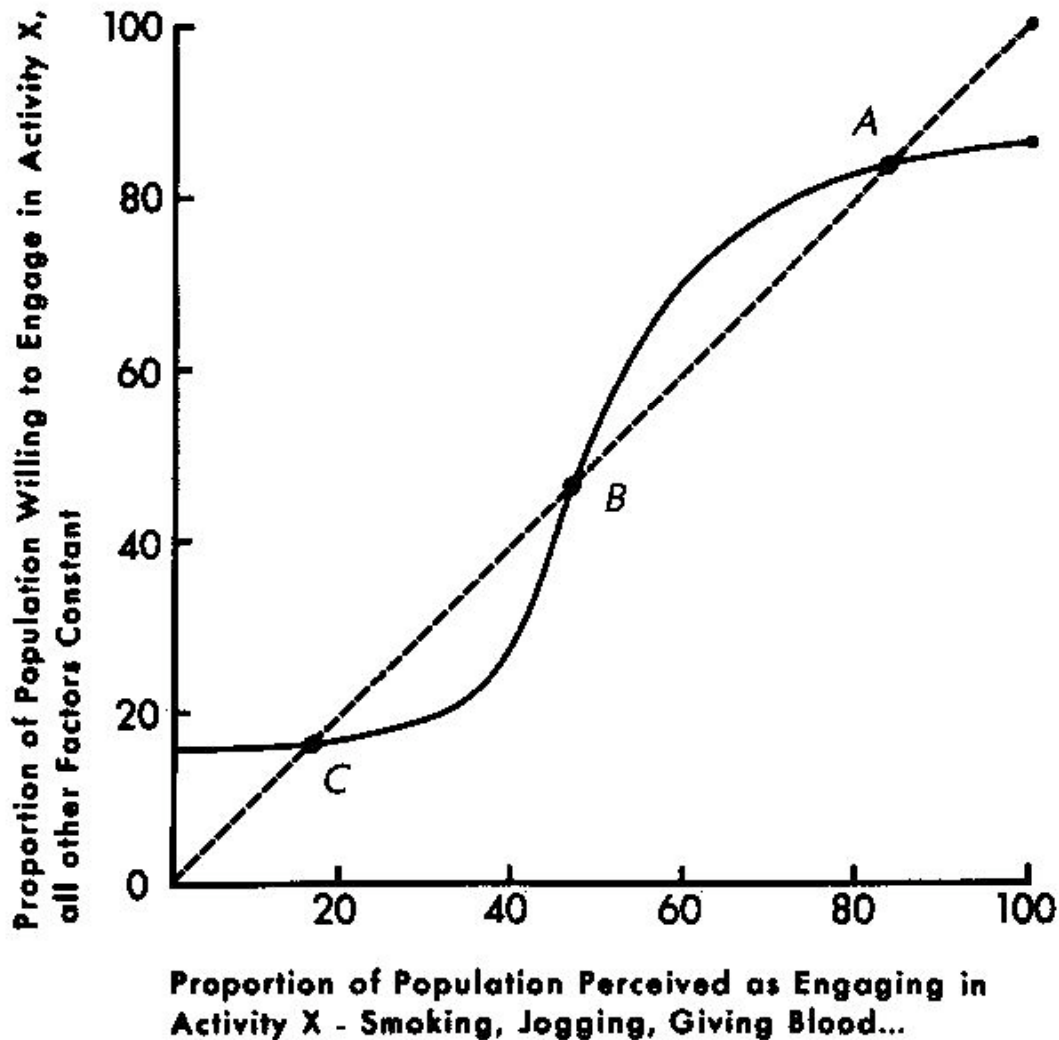
COMMUNICABLE BEHAVIOUR: AN EPIDEMIOLOGY OF LIFESTYLES?

A more sophisticated reinterpretation of the externality argument, however, goes beyond the individualistic postulates of conventional economic theory to observe that (as every advertiser knows) individual lifestyle decisions -- smoking, drinking, jogging, dieting -- are not taken in isolation. Each individual's lifestyle becomes part of the environment which affects the decisions of others. In part the interaction is psychological -- a herd instinct. Most people find it uncomfortable in and of itself to be the only one doing or not doing X. But there are also scale or threshold effects on the costs of particular activities. It is tough being the only cyclist in a city, for reasons which are quite directly physical, not psychological. When the baby boom reached middle age, the quality of jogging shoes rose spectacularly in response to the new market. When a large number of a firm's employees become fitness-conscious, it is worthwhile to provide centralized facilities.

Figure 12-1 (from Schelling 1978) depicts this process. The horizontal scale measures the proportion of a population perceived to engage in a particular activity; the vertical measures the proportion willing to do so. If no one else is doing it, a few rugged individualists are still willing to stand out of the crowd. As the activity becomes more popular, "bandwagon" effects develop, and a larger and larger proportion choose to participate. But at the other end of the scale, we find the equally rugged individualists who will not do X even if the whole rest of the world does.

FIGURE 12-1

A Representation of Interactions Among Members of a Community in Willingness to Engage in Particular Activities



Equilibrium is established where the S-curve cuts the diagonal -- desired participation equals actual.⁵ But, as shown in Figure 12-1, there may be several such equilibria, depending on how people respond to others' behaviour. Moreover some are unstable -- the equilibrium at *B*, for example -- and any shift in perceptions will cause a change in behaviour which is self-reinforcing. The *A* and *C* equilibria are stable; a shift in either behaviour or perception is self-damping. But high and low levels of participation are equally stable and equally the result of "individual choice." Private advertising campaigns often focus on shifting perceptions of participation -- most people, or beautiful people, or important people, people like you or whom

you would wish to be like, are doing X -- in order to sell particular products. Public prevention programs may likewise attempt to stimulate "healthy epidemics" -- shifts from high to low equilibrium levels of use of products which are harmful to health. The creation of smoke-free areas, for example, serves not only to protect the non-smoker from the smoker, but also to symbolize and reinforce the anti-social nature of smoking. Publicity campaigns serve not only to convey information about hazards -- which after all are widely known -- but to shift people's perceptions of what others do or approve. The tendency of smoking behaviour to begin among children in response to media and peer group pressure, and then to become addictive, emphasizes the inadequacy of the individual choice framework and the significance of external effects.⁶

But the informational aspect of prevention cannot be wholly discounted. The extent of current information as to the links between illness and either individual behaviour or environmental characteristics is woefully inadequate even among specialists in such fields, and *a fortiori* among the population as a whole. A public role can thus be justified both in supporting or conducting additional research in this field, and in disseminating information on the positive and negative effects of different activities or products. If the benefits of such research and dissemination could be captured by private individuals or firms in the process of selling particular products or services, then the public role might be redundant. In practice, however, prevention seems more often to involve *refraining* from use of particular goods and services. Moreover, the problems of imperfect information severely limit the effectiveness of the market response. The strong positive demand, willingness to pay, for prevention in a broad sense is indicated by the rapid growth in the last decade of sales of health and fitness products. But while in some cases (*e.g.*, jogging shoes) there has been a dramatic increase in the quality and quantity of technical and performance information available to users, and corresponding product improvements,⁷ in others -- health food stores -- the result has been the promotion of consumption of peculiar chemicals and dietary supplements whose effects on health are at best questionable.

Indeed private markets, being neutral on issues of health, lend themselves as readily to the promotion of unhealthy products and associated lifestyles, tobacco being the leading example. If a principal justification for public preventive activity is the need to remedy imperfect information and enable people to make more informed judgements as to the consequences of their actions, how can this be reconciled with the continued promotion of pathogenic products? Obviously, health promotion (and consistent policy) give way to political feasibility. The illogicality of giving tax deductions for the promotional expenses of tobacco companies and subsidies to tobacco farmers while funding public anti-smoking campaigns illustrates the fundamental importance of defining the standpoint in any process of program evaluation -- *cui bono*? More generally, the pathological agents who are the targets of preventive activity are commonly people and organizations with political influence, politically potent pathogens (Evans 1982a). The micro-organism or cell which is the target of clinical interventions may have a variety of weapons, but at least it does not advertise on its own behalf, vote, or make campaign contributions.

FROM LIFESTYLE POLICY TO HEALTH CARE EXPENDITURE: LONG CHAIN, WEAK LINKS

The difficulty of evaluating programs, or of making policy, in the area of prevention may be shown by considering explicitly the steps in the argument which lead from exhortation to

expenditure. It is frequently suggested that a particular program -- an anti-smoking campaign, a seatbelt law, dietary education in school programs -- will lead to reductions in health care expenditure, and thus pay for itself, saving money as well as improving health. Advocates of expanded preventive services often point to the present very large expenditures on cure or care, relative to prevention, as an irrational allocation of effort -- an ounce of prevention being worth a pound of cure.

The links in the chain run from a policy -- advertisements, regulations, changes in physician fee schedules to make preventive interventions more rewarding, to a change in behaviour -- people stop smoking, exercise, buckle up, eat less fats, see their physicians for check-ups, etc. This behavioural change is then supposed to lead to better health, lower rates of morbidity and/or mortality, which in turn are expected to lead to reduced health care utilization. And that, finally, should lower health care costs. Schematically:

Figure 12-2



Isolating each separate step of the process demonstrates the number of causal links which are necessary to support the argument for prevention as a cost-saving strategy, and the points at which it is vulnerable to attack.

Policy interventions include information or education campaigns, regulations, and taxes or subsidies on particular activities. It appears that when a form of unsafe behaviour can be narrowly targeted, regulations and tax policies are effective in modifying behaviour. When regulations on seat-belt use and driving speeds, or immunization requirements for school entry, are established and enforced, behaviour responds in a significant and measurable way. Similarly, taxation of tobacco and alcohol appears to have measurable and substantial effects on use levels. Determining the effects of educational or facilitational campaigns, however, is much more difficult. Dissemination of information does not necessarily induce behavioural change, and from the point of view of the health care system, knowledge enhancement is not an end in itself.

Nor does participation in particular activities, such as fitness programs, necessarily serve as a measure of behavioural change. To determine that an employee fitness program has changed behaviour one must know the pre-program activity levels among participants and non-participants, and the amount by which each changed when the program was introduced. To attempt to infer program effect by comparing activity levels of participants and non-participants would probably result in a serious upward bias in assumed effect -- participants being almost certainly more active to begin with.⁸ Determining the effects, if any, of a policy intervention requires a certain amount of sophistication in evaluation design -- simple comparisons can be seriously misleading for a variety of reasons.

If it is established that policies are available which can modify behaviour in desired ways, it is yet another matter to show that the behaviour will in fact have the desired effects on health status. Again, in narrowly defined areas like driving behaviour or tobacco use, the linkage from behaviour to health status is well defined. Speed kills, seat-belts save lives, smoking causes cancer, and water fluoridation reduces tooth decay -- uncertainty remains only out on the lunatic fringe. But the linkages from exercise or stress to heart disease are at best observed correlations whose causal significance is still questionable. It may well be true that physically active people

have less heart disease than inactive, but then sick people do not exercise much. A relationship between dietary cholesterol and heart disease has similarly been inferred for many years from correlations and speculation; not until early 1984 did experimental findings confirm the causal linkage.

The same problem of establishing causality applies to the dietary, occupational, or environmental correlates of cancer. And trying to control morbidity or mortality by influencing a risk factor which is not causally connected is like trying to cool a room by putting ice cubes on the thermometer.⁹ Unfortunately, the present state of knowledge appears to be insufficient to establish hard and fast causal links between behaviour and the major lifestyle-related sources of illness, cholesterol being a very recent exception. In this situation, the payoff to preventive programs is inevitably rather difficult to establish.

PREVENTIVE SERVICES IN THE CLINICAL SETTING: PROBLEMS IN ASSESSING EFFICACY

Clinically focussed preventive programs are subject to the same sort of problems. If one thinks of willingness to undergo screening procedures as a form of behaviour which can be promoted by exhortation, subsidy, or regulation, then the linkage from that behaviour to health status is increasingly questioned. It seems now generally agreed that routine "annual physicals" for the population at large are ineffective, and that screening should be targeted only to pre-selected high-risk groups (Canada, Task Force on the Periodic Health Examination 1980).

There are numerous specific traps in the determination of efficacy of clinical preventive manoeuvres, arising from the incompleteness of the underlying knowledge base.¹⁰ Early detection of illness, in a pre-symptomatic stage, is preventive only if intervention is more effective or less costly at an early stage in the evolution of the condition. If the outcome of therapy is measured, however, by survival rates, after, say, two, five, or ten years from date of diagnosis, as is the case for cancer, then early detection may artificially raise measured survival rates by bringing forward the date of diagnosis even if the course of the disease is unaffected. The patient is actually made worse off, by spending longer in a "labelled" condition of pre-symptomatic illness, without gaining any corresponding benefit from therapy. In evaluating the benefits of early diagnosis and intervention, one must be sure that survival rates are measured from when the illness would have been identified in the absence of early detection, but this may not be known. If an early detection program is associated with increases in measured five-year survival rates, but also with an increase in population-wide prevalence of the condition, and no change in mortality rates, one must suspect that its influence is only on date of diagnosis.

A similar problem arises in the case of false positives. Few tests are absolutely precise in distinguishing pathological conditions; most are calibrated to trade off type I and type II errors. Some cases of "illness" will be missed, false negatives, and others wrongly diagnosed as ill, false positives. But if treatment is initiated for all positives, then the success rate will be contaminated by the false positives who will presumably show up as "cured." In the limit, a totally ineffective therapy will have a "success rate" equal to the ratio of false to total positives in the testing process which diagnosed the condition. If early detection programs lead to pre-symptomatic intervention, they may raise the proportion of false positives treated, and hence the "success rate," even if the therapy is totally ineffective. This problem is particularly likely when early detection identifies prior conditions -- "silent" gallstones, carcinoma *in situ* -- whose probability and time path of transition to a symptomatic condition are uncertain. The positives are not

exactly "false," but a large number of people may be "successfully treated" who would never have become symptomatic in any case. Again, the proper approach is to identify population-based experience and compare it with controls -- if the preventive program results in an increase in therapeutic intervention, an increased "success rate," but little or no change in morbidity or mortality in the target population, then the payoff to intervention may be illusory.

The point is not that clinical or other preventive interventions never "work" nor that reliable evaluation is impossible. Techniques of evaluation design and statistical analysis exist to surmount these problems, and to identify the critical gaps in knowledge. That is part of what epidemiology is about. But casual approaches to evaluating efficacy of preventive interventions are even more subject to bias and error than clinical interventions generally. The simple faith that "an ounce of prevention is worth a pound of cure" will be misleading if we do not in fact know how to prevent, and naive evaluation can easily hide our ignorance from us.

The uncertainties associated with links one and two in the chain in Figure 12-2 are sufficient to undermine a very popular argument for prevention, based on gross spending levels. It does *not* follow that just because much more is spent on cure or care than on prevention, therefore preventive activity should be expanded. The key questions are the marginal ones: What will be the payoff, in terms of health status, if more resources are added to either prevention or cure? What is the slope of the curve in Figure 1-3 with respect to these alternative activities? If preventive activities are ineffective, then no amount of spending on cure (or good intentions by their proponents) justifies their expansion.

Of course in practice the situation is more likely to be that such activities have not been *demonstrated* to be effective, but they might be. Under these circumstances, one must choose between investing resources in a program of uncertain payoff, or investing in research to make the payoff more certain. To this dilemma there is no general answer; presumably the choice depends on the urgency of the situation, the degree of uncertainty, and the cost and expected payoff of research. But as a generalization, "hoping to goodness" is likely to be expensive as well as theologically unsound.¹¹

HEALTH STATUS AND THE LIFE CYCLE OF CARE USE: DEAD MEN USE NO SERVICES

The marketing of preventive services as a method of reducing current levels of health care expenditure, however, also depends on the next two links in the Figure 12-2 chain. These are prone to failure for reasons unrelated to imperfect information, but rooted in the structure of the health care industry and in the human condition itself.

It is regrettably not true, in general, that improving health status lowers health care utilization -- it may or may not. The ambiguity is partly a result of the distinction made by Thomas (1971) between half-way and decisive technologies, and partly stems from the inevitability of the aging process itself. The development of polio vaccine was clearly an example of a decisive technology which both contributed to health status and lowered health care utilization. On the other hand, seat-belt use shifts the whole distribution of injuries so that some are alive who would otherwise be dead. It is a researchable question, whether the totality of health services used by accident victims is higher or lower when seat-belts are used; the answer cannot be given a priori. Similarly, hypertension screening and drug therapy create a population of regular users of physicians' services who would not otherwise be symptomatic or labelled "ill." The incidence of stroke may fall as a result, but whether utilization of health services by

stroke victims falls by enough to compensate for the increase in use by hypertensives under care -- keeping in mind that strokes may be fatal -- cannot be answered a priori.¹² (But it *can* be answered, as the seat belt question can, by analysis of utilization data).

Apart from the problem of identifying the impact of a successful preventive intervention on utilization patterns, by the target population, of care related to the intervention, there is also the issue of unrelated care. Everyone dies, and in our society, most people become aged. In the process, various biological systems fail. Medical, drug, and particularly hospital utilization rise sharply with age. Preventive interventions which lengthen life also increase the number of years spent in the high use period. The overweight smoker who dies on the street of a coronary at age sixty-two will generate no further utilization; the energetic jogger who lives to be ninety-two will probably require a substantial amount of maintenance care in the final thirty years.

Or she may not. It is possible that improvements in health status resulting from adoption of healthy lifestyles may extend the horizon of good health and reduce the needs of the elderly at any specific age. The well-preserved seventy-year-old of the future, if prevention works, may display the utilization patterns of today's fifty-year-old. But at present we do not know. The case can be argued either way (Fries 1980; Schneider and Brody 1983), and the jury is still out. It would therefore be unwise to assume away the possibility that preventive services, if effective, will actually increase per capita health care use. In any case, age-specific use rates among the elderly are now rising, not falling and have been for a number of years (Evans 1984) but factors other than health status may be at work.

SAVINGS AT WHOSE EXPENSE? THE INCOME-EXPENDITURE IDENTITY AGAIN

Finally, even if a preventive program is able to thread its way through the causal chain in Figure 12-2 and actually serve to lower health care utilization by its beneficiaries, it does not follow that health care costs will fall as well. As emphasized above, health care expenditures are by definition also the incomes earned from the health care sector. Thus a policy which is successful in limiting expenditures must also lower incomes -- again by definition. If British Columbia's compulsory seat-belt law were really to save \$120 million in health costs, as claimed when it was introduced, that \$120 million must be made up of reduced fee income of orthopods, fewer nurses hired, hospital beds closed -- resources either withdrawn from the production of health care, or paid less for their services. If neither of these happens, then no money is saved.

In practice, however, the influence of care providers over their own activity levels is such that resources do not go unemployed. The surgeons who are no longer re-assembling accident victims can instead replace arthritic joints; the nurses and beds are not left idle but are used to care for other types of patients. And if workload and incomes do fall in the fee-for-service sector, this translates into increased pressure in fee bargaining. Only if the preventive policy is matched with deliberate independent efforts to close beds, reduce hospital budgetary allocations, and limit the supply of medical practitioners will its effects, if any, on utilization result in expenditure savings.

Of course the redeployment of resources may itself be a gain, depending on what is believed to be the payoff to additional resources in health care. But if public policy generally is struggling with a perceived oversupply of physicians and hospital beds, it is clearly inconsistent to argue for a preventive policy on the grounds that it will serve to make more health care services available for other purposes.

This is *not* to say that such a policy would be valueless. An effective preventive policy which contributes to improved health can be evaluated on that basis alone. But it can easily be misleading to go beyond such health benefits and argue that, in addition, the policy will ease expenditure pressure on hard-pressed governments. When the savings fail to materialize, or are eaten up elsewhere, the resulting loss of credibility may jeopardize all preventive initiatives, effective and ineffective alike.

This possibility is of particular concern in times of fiscal restraint, when it has become increasingly tempting for advocates of health care programs of all types to try to market them as cost-saving, on the basis of evaluation analyses ranging from inadequate on down. New and expensive drugs and high-technology diagnostic and therapeutic interventions are always guaranteed, along with their contributions to patient well-being, to avert other misfortunes or supersede other expensive procedures, such that despite their direct costs, overall costs will fall. Preventive programs often mount the same bandwagon.

PREVENTING POVERTY AMONG PROVIDERS?

Yet a cynical observer might suspect the reverse intention. Recalling the identity of health expenditures and health incomes, much so-called "preventive" activity may be viewed as an effort to expand the scope and intensity of health care interventions to include the currently healthy population. Curative and carative interventions are to some extent limited by the presence of pathology in a population, and their effectiveness can be judged relative to that presence. But prevention is a completely open field; there is no limit on the amount of potential evil which one might hope to avert. And after the fact it is very difficult -- impossible at the level of the individual -- to determine if the prevention was effective.¹³

Thus prevention serves as an excellent way of absorbing the energies of an expanding health labour force in activities which probably do no harm, may do some good, and offer the psychological rewards of the celebration of wellness. Moreover, since the range of alleged, and possibly valid, preventive techniques extends well beyond the conventional health disciplines, prevention creates a point of entry for other practitioners into the very favourable regulation and reimbursement structure enjoyed by health care providers. Not surprisingly, reimbursers and currently approved providers find themselves allied in opposition to this expansion in the supply of health care incomes.

But the preventive interventions which *have* been proven effective in improving health are largely or wholly outside the clinical field. Tax and regulatory policy have been shown to be effective (and cheap) in controlling smoking and drinking behaviour and seat-belt use; efforts to influence such behaviour in a clinical setting have yet to be shown effective and are certainly more expensive. Fluoridation of water supplies is very effective and cheap; prevention in the dental chair is of much less effect and relatively expensive. Mass immunization campaigns are cheaper and more effective than initiatives by private physicians. There may, of course, be exceptions to this generalization -- hypertension control and prenatal care being possible examples of effective forms of clinical prevention. (More generally, see Morgan 1977.) But there does appear to be a trade-off, both in principle and in practice, between the interests of the community in cost-effective prevention, and those of health providers in expanding the markets for their services. There is a real danger that the health care industry may find itself allied with other interests in the community which are threatened by effective preventive interventions, in the promotion of expensive and ineffective clinical alternatives. As an example, responding to

lung cancer by more cancer research and by radiological and surgical interventions, rather than by controlling smoking, serves the economic interests of scientists, clinicians, and tobacco producers. But it is an expensive way of not achieving the allegedly desired result.

On the other hand, the inadequacy of the economic self-interest model in explaining professional behaviour is clearly shown by the universal support of dentists for fluoridation of community water supplies, a highly effective preventive intervention which directly substitutes for clinical services, both restorative and preventive.¹⁴

HEALTH CARE RESEARCH: WHAT IS THE PUBLIC INTEREST, AND HOW IS IT BEST PURSUED?

Health research, like prevention, displays two faces. On the one hand, it represents a form of social or private investment in the acquisition of new knowledge, which can improve the effectiveness and/or efficiency of the health care delivery system. In terms of Figure 1-3, the curve linking health status to resource inputs should be shifted upwards and to the left by successful research. On the other hand, research and the technological innovations it makes possible are frequently criticized as contributing to the escalation of health care costs -- presumably by shifting society out along the relatively flat section of Figure 1-3, or even into the negative slope -- and reducing the quality of care by dehumanizing and mechanizing it -- losing the curative dimension -- without corresponding gains in curative effectiveness.

In principle, research activity can be fitted into the cost-benefit framework of equation 11-1, and there are some examples of studies which measure the costs of and payoff to particular research efforts. One can add up, suitably priced and discounted, all the resources devoted to achieving a particular piece or collection of new knowledge, and compare these with the benefits in terms of reductions in other health expenditures, increases in productivity of those spared, or cured of, particular illnesses, and the overall value of lives extended in quantity and improved in quality.¹⁵ These benefits are available to the community as a whole, assuming the results of research are widely disseminated, and private individuals or groups undertaking research will thus bear all the costs but reap only part, perhaps a small part, of the benefits. Research will, in general, be underfunded unless the public sector provides some sort of support.

There is substantial choice, however, as to the form of support; and the relative effectiveness of different approaches, for different types of research, is itself an under-researched topic. The state can conduct research itself, as through the National Research Council or the Defense Research Board or agencies in particular Ministries. It can make grants to not-for-profit agencies or their employees -- universities, research institutions, hospitals -- as does the Medical Research Council or the National Health Research and Development Program. It can let research contracts to private for-profit firms selling research services. It can grant special tax treatment or other forms of public subsidy to the exploitation of new research information. Or it can grant patent protection to private firms, or give them privileged access to government purchasing programs, thus increasing the commercial value of their in-house research.

The inclusion of patents in the spectrum of public policies in support of health care research may seem somewhat unusual. From an economic perspective, however, a patent right is an asset, with an economic value, conferred by the state. The state creates and maintains, by coercion if necessary, a monopoly position for the patentee which enables him to "tax" purchasers of the patented product by charging them a price above production cost, and above what a competitive

market would establish. It is thus like an agricultural marketing board, a delegation of the state's taxing power to a private entity for its own benefit.

The purpose of such delegation is twofold. First, in the absence of patent protection innovators would have an incentive, where possible, to keep secret the results of their research. The resources thus devoted to security, and by their competitors to industrial espionage, are a deadweight loss to society as a whole. Moreover, insofar as innovators are successful in maintaining security, they retard the dissemination of new knowledge and lower the rate of technical progress generally. But if security is weak or impossible, then successful research conveys no commercial advantage because it can readily be copied by competitors, and no one firm will find it worthwhile to undertake.¹⁶ The level of investment in research will be too low. Patent policy thus becomes a delicate balance between providing too much protection, such that the monopoly power granted to innovators imposes excessive allocative and distributional distortions on the relevant industry, and not enough, such that the rate of innovation is too slow and/or the level of effort in industrial security is too high.¹⁷ But it parallels the political questions of how much to spend on in-house, or grant to "out-house" research; in each case the state must, by using the direct or indirect taxing power, balance present resources used up or benefits foregone against future benefits gained.

CONSISTENT MANIPULATION OF UNKNOWN QUANTITIES: THE ECONOMIC ANALYSIS OF RESEARCH

The difficulty with cost-benefit or cost-effectiveness analysis in this context is that, while it can in some cases be done after the fact, it is notoriously difficult to do in advance. It is painfully true that "if we knew what we were doing, it wouldn't be research." Of course, in some areas the nature of the objective may be well specified. The substantial amount of research carried out by private firms is generally focussed on a particular product -- a drug or machine -- whose desired effects or operating characteristics are relatively well defined. There may, however, be substantial uncertainty about the time and effort required to develop them. For more basic research the eventual payoff, if any, may be very difficult to predict and may occur in fields apparently quite unrelated to what the researcher or supporter had in mind. It is in just these areas, where some participants feel the highest long-run benefits may lie, that the necessity for public support is greatest. But since the B values of equation 11-1 are almost totally unspecified, if not unspecifiable, the potential contribution of economic analysis is distinctly limited. And its limited contribution is primarily negative, in that it can clear away some of the fuzzy thinking surrounding research policy and reveal the problem in its full intractability.

In the first place, economic analysis emphasizes, as always, the essentially marginal nature of the policy problem. The issue is not whether a society should support health-related research at all, but whether the \$327.1 million reported as spent in Canada in 1982 was too much or too little.¹⁸ And was it spent in the areas of greatest probable payoff? The rhetoric of technical progress or regress too often seems to be devoted to maintaining that research is a GOOD THING, which no doubt it is, but so are many other things. The whole of the GNP should not be spent on health research, nor even half of it, but what criteria can one apply to set a level? The judgements of professional researchers about appropriate funding levels overall are obviously unreliable, both because researchers are far from disinterested, and because they lack either the professional expertise or the political legitimacy for making such balancing decisions. They may,

however, on the basis of their expertise advise as to the outcomes to be expected from different effort levels, and on the mix of activities or appropriate targets.

Secondly, there is no automatic linkage between appropriate research levels and the economic or human magnitude of a health problem. The efforts by the Ontario Council of Health (Fraser *et al.* 1976) were an interesting exercise, but provide little guidance for recommendations on research funding. As in the case of prevention, the fact that much is spent on curative or carative activities does not, in and of itself, indicate that more should be spent on research (or, for that matter, less). "Magic ratios" of 1 percent or 2 percent or whatever of health spending to be devoted to health research, have unfortunately no rational basis beyond the aspirations of researchers. The key question is not simply, "How big is a particular problem?" but "What probability is there that a particular research effort will actually contribute to the solution of a problem large enough to justify the effort?"

Estimates of the costs, direct or indirect, associated with a particular class of problems may have negative significance -- a small problem obviously does not warrant a large research effort. But the converse need not hold. For example, the finding by the Ontario Council of Health that diseases of the teeth and supporting structures generate a relatively large direct economic burden is not an argument for more research in oral biology. As noted above (chapter 7), a large proportion of this cost is a result of technical inefficiency in the delivery of dental service, overutilization of dentists instead of dental nurses to carry out restorations, a result not of lack of knowledge but of professional self-protection. An additional component results from incomplete application of well-established efficacious prevention -- community water fluoridation. If new knowledge cannot be expected to lower costs, whatever other benefits it may entail, then high costs cannot be an argument for more investment in research.

"FREE RIDING" VERSUS NATIONAL PRIDE: OR DR. BANTING, MEET MR. PODBORSKI

Nor is a higher level of research effort in other countries an argument for more research effort in Canada, quite the contrary. The more other countries spend on research in a particular area, *ceteris paribus*, the less we should spend in Canada, because the more likely we are to benefit from others' work, and to find that our own is duplicated. We should concentrate on areas that others have left alone. Of course, the *ceteris* are not *paribus*; it may be that others have left an area alone because they judge the expected payoff to research there to be small. But for a wholly self-interested small country in a large world, the optimum research strategy may be none at all -- better to ride free on the product of others' efforts.

There are, of course, other considerations. "Free riding" carries a mildly offensive connotation, however rational it may be. Canadians' self-image, of themselves and their country, may be enhanced by a feeling of doing their share on the world scene. Supporting research efforts, and perhaps Nobel Prize winners, is a source of collective satisfaction akin to supporting Olympic medallists or downhill racers, or beating the Russians at hockey. Most Canadians take vicarious pleasure in, and are willing to pay something to support, such achievements. But there is no necessary reason why such support should be particularly for health research.

One can also make a sort of "infant industry" argument, that support for research can help to develop high technology industries in Canada, with consequent benefits in jobs, skills, earnings, and profits. But this is not an argument particularly for health research, and insofar as it does bear on health research, it is only as a means to production and sales, not as a way of improving

health care. Assisting the development of Connaught Labs as a multinational pharmaceutical firm, for example, may or may not be a proper part of Canadian industrial policy, depending on the expected net payoffs, but it has nothing at all to do with health policy. In fact, the effects may be negative if the Canadian health care system is viewed as a protected market for the domestic development of exportable health care products whose health benefits are at best marginal. Industrial subsidies become hidden in health budgets. The profitability and growth of health care suppliers and the cost-effectiveness of the health care system may, as noted in chapter 10, be in direct conflict.

On the other hand, new knowledge does not always arrive neatly packaged on the shelf. Domestic skills may be needed to adapt new information, and these skills may be maintained only within a domestic research program. When a new drug is made available in Canada, its effective utilization does not depend on whether it was developed in Toronto, New Jersey, Belgium, or outer space. But new therapeutic techniques or management systems may require localized expertise for their adoption. The hard-nosed free-rider policy may lead to a country's being slower to adapt new knowledge because its providers are unaware of its existence or significance, or simply have to learn how to use it.

And the use of research, the delivery of care, may not be separable from the research process itself. Opportunities to advance human knowledge, or to satisfy their curiosity, form part of the professional self-expression of some at least of the best clinicians. Just as high quality university teaching requires the presence of active researchers, so it is argued that the maintenance of clinical standards, striving for excellence, may depend on the proximity of an active research environment. Otherwise professionals will migrate to foreign research centres, or to the ski hill.

Finally, health-related research spans a continuum from the "hard end" -- physics, chemistry, or biochemistry -- through the biological sciences to the social and managerial disciplines. Results from the former know no national boundaries, but the latter are culture and system dependent. Health care system research in the United States or Sweden may have some implications for Canada, or it may be completely irrelevant, depending on the extent of similarity of the institutions or behaviour studied. We each have to do our own research on our own health care systems, if the results are to be helpful in addressing our problems.

These are all rather general considerations, which do not provide much concrete guidance to issues of research funding and priorities.¹⁹ And that is probably inevitable. If the expected benefits of research are intangible and highly uncertain, the techniques of economics are singularly weak. At the very least, however, one can ask advocates of increased research funding -- whether in the form of grants, favourable tax treatment, or increased patent protection -- to come up with some plausible estimate of the resulting benefits.²⁰

And it is certainly important, as in any public or private expenditure program, to scrutinize carefully the results of such efforts. (They may not be what was expected, or promised, but there should be *some* results!) In the end, though, it is undoubtedly worth supporting a certain amount of "pure" research, whose expected payoff is totally beyond quantification. How much, is essentially a political decision. Economics may help to clarify some of the issues involved, but cannot add much to the final answer. How much was Expo '67 worth, or the Montreal Olympics? Our collective judgements on such matters are made, albeit rather imperfectly, through the political process.

THE PROFITS ARE IN "HALF-WAY" TECHNOLOGY: PROBLEMS OF MOTIVE

In the special context of health economics, however, there is an additional consideration. Thus far, as in most discussions of research activity, we have implicitly assumed that new knowledge was always a positive good. If so, the payoff to research is always non-negative. A good program produces a lot of new knowledge for the money, a bad program little or none. But one cannot be worse off, as unhelpful new information can always be ignored.²¹ Yet in health care, it may be that valid research can have harmful consequences. In the first place, as noted in chapter 10, the economic incentives which bear on the choice of research programs are predominantly towards cost escalation.

Consider Lewis Thomas's (1971) three-stage characterization of health care technology. The first stage, palliative care for an illness whose course cannot be influenced, is usually cheap and ineffective (though becoming less cheap). As knowledge progresses, we learn to deal with the consequences of the illness -- iron lungs, organ transplants, renal dialysis -- to keep the patient alive and perhaps functioning better. This "half-way technology" stage is partially effective and very expensive. Finally, when knowledge advances to the point that the disease process itself is fully understood and controllable -- the real high technology stage -- the decisive intervention is again cheap but now effective.

Thomas's argument leads toward more "basic" research in biology and biochemistry, in the hope of reaching stage three, rather than "applied" research, which focusses on specific problems at the stage two level. The argument is attractive, though it does not tell us how much basic research to support -- or who to ask for the answer. From a private economic perspective, however, stage two technology is the profitable level to address. Expensive half-way technology yields sales and profits. Decisive technology removes the problem -- and the market. Indeed, it is a positive (commercial) advantage that, for example, micro-organisms become resistant to antibiotics -- there is always a market for new drugs.

This rather macabre perspective should not be interpreted as imputing malevolent motives to individuals or firms; it merely addresses the obvious. Profits are earned by selling things, not by not selling them. It follows from this that research directed to producing high sales will have a large expected payoff, and that profit-oriented firms will, quite rationally, tend to support that sort of research. Moreover, insofar as the adoption of new technology by providers of health care is biased toward that which is complementary to their skills, enhancing their sales potential, and away from that which substitutes for their services and is cost-reducing, this will further bias the research process toward the development of cost-enhancing technology.

This points up the weakness in the argument that technology per se is neutral and that problems arise only from the way in which health care providers use new information. If technological advance is associated with cost escalation and ineffective utilization, that is the fault of the workman, not the tool. But the argument is too simplistic. Insofar as research is conducted by for-profit firms, it will be focussed differentially on those developments which are most likely to be overutilized -- overutilization translates into sales volume -- and the promotion of products embodying new knowledge will similarly encourage use separately from efficacy. Hence, the frequent observation of unquestionable technological breakthroughs which improve the effectiveness of care, and in a particular context lower its cost, but which tend to be used far beyond their range of demonstrated efficacy, with corresponding cost-escalation effects. The research process itself is not, cannot be, neutral; much less the process of diffusion of new knowledge into practice.

AFTER SUCH KNOWLEDGE, WHAT FORGIVENESS?

Finally, on a still more difficult level, health care is not motivated solely by efficacy considerations. The sense of having "done everything possible" for the patient is itself of value to providers and relatives, if not necessarily to the patient herself. The significance of the desperation reaction in technological diffusion (Warner 1975) may be partly based on the same feeling -- if the outcome was not satisfactory, at least nothing was left undone.

In this context, new knowledge may simply add to the range of things which should not be left undone. The tendency for new diagnostic techniques to be applied along with those they were expected to replace appears to be an example of this phenomenon. And it is not merely a matter of provider uncertainty. For the patient and the wider society as well, unavoidable evils are easier to bear and less productive of distress or outrage than those which might possibly have been avoided -- if someone had done something.

In a fully rational world, of course, people would make judgements about expected payoffs, would decide who should and who should not be treated, and would then go home and sleep well (or suffer in silence). In our world, however, the development of new technology with very high costs and low but non-zero benefits may well make people worse off, in that now more decisions to withhold treatment must be made -- or else costs must be allowed to escalate. In terms of Figure 1-3, an increase in diagnostic or therapeutic knowledge which increases the slope of the curve from zero to some very small positive value is by no means an obvious gain in well-being. Such new knowledge, while privately profitable, may carry negative externalities -- be a public "bad." But does that mean we should try to suppress it, assuming that we could?

NOTES

¹ This cross-national issue arises to a much lesser degree in prevention, but Canadians did benefit from the WHO program to eradicate smallpox. The risk of contracting the disease fell from very low to zero, and preventive activities in Canada, such as immunization, ceased to be necessary. Successful prevention elsewhere in the world thus yielded domestic benefits which could not have been achieved with any amount of local effort.

² Compulsion is paternalistic to the extent that it requires people to do what others think is good for *them*, as well as for their neighbours. This would include, for example, a requirement that all female children entering high school be immunized against rubella. But non-paternalistic compulsion would cover quarantine or other specific disabilities imposed on the ill for the protection of others. In terms of equation 11-1, public subsidy or provision seeks to lower the private costs of socially beneficial activity while relying on private decisions; paternalistic regulation responds to different shadow prices, P and V , on the consequences of the action and compels the individual to act as if she held these different values; non-paternalistic compulsion may overtly recognize that the preventive behaviour is undesirable for the actor but require it in the interest of others.

³ In theory, an individual suffering damage from the actions of another can claim redress through the courts. An ideal and castle tort system would internalize external effects and thus remedy a major source of failure of private market systems. It would also eliminate a principal justification for public intervention. In practice, however, the problems of establishing liability, both fact and quantum, in addition to the sheer administrative inadequacies of the tort system (Dunlop 1982), make this prospect as illusory as the parallel interest in professional quality assurance through malpractice insurance rather than licensure.

⁴ In a hypothetical world of perfect information, insurance contracts would be written such that each person's premium reflected all aspects of her behaviour bearing on risk. All costs would then be internalized; A's premium would fully cover the (expected) costs of A's behaviour and B would be unaffected. But the costs of monitoring plus the inadequacy of epidemiological data on risk make this hypothetical situation irrelevant in practice. Private,

competitive insurance systems do introduce more premium differentiation than public, on an experience-rated basis, but this links costs to group health status, not individual behaviour.

⁵ Since desired participation depends on perceived participation, systematic distortions of perception which do not die out lead to somewhat more complicated equilibration processes.

⁶ Apart from the obvious evidence of advertising campaigns themselves, anti-smoking campaigns among school children have found that non-smokers and smokers together greatly overestimate the proportion of children in their schools who smoke. Thus the perception of peer pressure is accentuated. ("Come on, *everybody's* doing it!") And as Figure 12-1 emphasizes, it is *perceived* collective behaviour which matters. Actual behaviour generates external effects of this form only insofar as it affects others' perceptions of prevalence. Pressure to keep certain forms of behaviour unobtrusive or even secret may therefore shift the curve in 12-1 and affect actual patterns. "Coming out of the closet" can affect actual patterns of behaviour, as well as perceptions.

⁷ Though the most widely used source of such objective information ceased to publish it between 1981 and 1982, and reverted to a "popularity poll." Magazines must respond to the interests of their advertisers, which may call into serious question the potential for information generation and dissemination by private, for-profit firms in a competitive marketplace. Magazines like *Consumer Reports* and *Canadian Consumer* accept no advertising and are non-profit -- and one to a country.

⁸ The same criticism applies to reported savings of sick time and absentee rates generally, or of higher employee productivity and morale, from fitness programs. It seems very probable that such programs differentially recruit the healthier, fitter, more company-minded employees. Ideally, one would like to assign people randomly to participation and control groups and compare the performance of each; failing this, paired before-and-after comparison would probably be conclusive. In either case, though, the nonparticipation group might conceivably be contaminated by "halo" effects.

⁹ If the thermometer happens to be part of a thermostat, the consequences are still more awkward.

¹⁰ This section in particular leans heavily on the work of Sackett and others at the Department of Clinical Epidemiology and Biostatistics, McMaster University, who are however in no way implicated in its inadequacies.

¹¹ It is sometimes argued that it is unfair to subject preventive programs to rigorous standards of evaluation which have not been met by the vast majority of curative and carative interventions. Yet the ultimate objective is surely not one of fairness to different providers or balance among unproven techniques, but of benefit to patients and taxpayers. If resources are being wasted in one area, it is not obvious that waste in others is thereby justified. On the other hand, if some degree of waste is inevitable, perhaps a balanced service offering does require symmetry of evaluation standards. Certainly the status quo should not be exempt from scrutiny.

¹² The hypertension case adds an extra complication. Case-finding *labels* people as "sick," leading to real psychological costs for the patient as well as, in some cases, behavioural changes with measurable (negative) economic effects. But not all patients will comply with therapy. Those who do not are clearly made *worse* off by the case-finding intervention (Haynes *et al.* 1978). The same is true for patients labelled by a screening process, for whom therapy turns out to be ineffective (earlier detection of an inoperable cancer).

¹³ Most of those who undergo a preventive intervention would not suffer the evil in any case, and even if the intervention is effective, some of its subjects may still fall ill. Efficacy can only be determined at a population level, often over a long period of time.

¹⁴ It is conceivable that, by postponing or preventing tooth mortality, fluoridation might increase dental service utilization on a lifetime basis -- the edentulous have little need for care. But to suggest that such expectations explain professional support for fluoridation seems not only cynical, but silly.

¹⁵ Weisbrod's (1971) work on polio research demonstrated the evaluation of a clear-cut research success, whose costs and benefits could be identified and compared.

¹⁶ In some fastmoving fields copying may be too slow, and market power depends on speed of innovation. In others, production-specific "know-how" may be difficult to transfer. In these situations, patent policy may be relatively unimportant.

¹⁷ In a static equilibrium world, the optimal patent life is zero, as is the technical progress rate. Patent policy presupposes a trade-off in terms of welfare losses, departures from Pareto optimality, at each point in time, in return for a faster growth rate over time. Optimizing across this trade-off is a social, not an individual, problem since the requisite state power is collectively held.

¹⁸ An additional amount was spent through the prices of health products, particularly prescription drugs and some hospital equipment, but this research was almost all conducted outside Canada and is not an object of Canadian policy.

¹⁹ Though the easier importability from other countries of research results at the "hard end" would suggest that rational policy for a small country in a large world is to concentrate on social and managerial research -- health systems and delivery -- not basic science. In particular it casts doubt on the wisdom of devoting resources, either directly or through the patent system, to encouraging domestic pharmaceutical research by multinationals; this is the easiest of all research to import.

²⁰ The suggestion, for example, that Canada modify or end its program of compulsory licensure of the importation of patented drugs, in return for increased research spending by drug companies in Canada, represents a transfer of funds from Canadians (and their governments) as drug buyers to pharmaceutical manufacturers for research purposes. Whether this would lead to increased research worldwide is unclear, though presumably Canadians would regard increased domestic employment as a benefit. But the results of such research, though supported by Canadians through higher prescription prices, would remain the property of the companies concerned. Canadians would have no privileged access to the information generated by such research, whether or not carried out in Canada, even though they had effectively funded it. The proposal would in essence be equivalent to a job-creation program run by private companies with federal grants raised by a tax on particular pharmaceuticals, without any effective linkages between the size of the grants and the number of jobs created.

A policy of selective public assistance to pharmaceutical research as part of a long-run strategy of "Canadianization" of pharmaceutical production, or at least the development of one or two world-scale Canadian-based producers (Gordon and Fowler 1981) raises more interesting issues both pro and con, but they are beyond the scope of this work.

²¹ Of course, research may lead to erroneous findings. The hard sciences have relatively effective ways of cleaning out bogus results fairly quickly. But at the other end of the spectrum, misleading results from, for example, bad economic research can have powerful negative effects. Most people, including economists, would agree with this statement, but would disagree over what was bad research -- which makes the point.