Should physicians be gatekeepers of medical resources?

Milton C Weinstein

Harvard University, Boston, Massachusetts, USA

Abstract

Physicians have an ethical responsibility to their patients to offer the best available medical care. This responsibility conflicts with their role as gatekeepers of the limited health care resources available for all patients collectively. It is ethically untenable to expect doctors to face this trade-off during each patient encounter; the physician cannot be expected to compromise the well-being of the patient in the office in favour of anonymous patients elsewhere. Hence, as in other domains of public policy where individual and collective interests conflict, some form of collective solution is required. Collective solutions may take the form of placing explicit resource constraints on resources available to physicians, or clinical practice guidelines that recognize cost-effective care as acceptable. Such solutions will be politically and ethically sustainable only if patients as citizens of the larger population accept the need for rationing of limited resources in health care.

(Keywords: Resource allocation; physician-patient relationship; cost-effectiveness)

The medical commons

Resources available for health care are finite. This means that it is impossible for physicians collectively to offer all technologically feasible and clinically beneficial medical services to all patients. Individually, though, physicians practise medicine under a basic ethical tenet which compels them to do whatever is in their power to help their patients. Herein lies one of the fundamental ethical issues in modern medicine: how can physicians fulfill their moral obligations as fiduciary agents for individual patients while being responsible stewards of the finite pool of resources?

In a 1975 article in the New England Journal of Medicine, Howard Hiatt likened the situation in medical care to a parable described by Garrett Hardin in a classic article, entitled The tragedy of the commons. According to Hiatt’s adaptation of Hardin’s parable, physicians are like herdsmen who used to feed their cattle (patients) on a common pasture. Acting in their own interests, the herdsmen could allow their cattle to feed on the land without limit, as long as their numbers and appetites were small compared to the resources on the land. As the number and appetites of the cattle grew to the point where, collectively, their wants exceeded the capacity of the common resource, the desires of the herdsmen to extract the maximal nutrition for their cattle led to overgrazing. At first, the less aggressive cattle failed to get adequate nutrition and died. Later, herdsmen were forced out of business. And, in the end, the rich pasture turned into an overgrazed wasteland.

Although Hardin’s essay was written in the context of population growth, Hiatt saw its relevance to health care. Physicians, each acting in the best interests of their own patients, collectively reach the limits of health care resources, with the result that access to care and quality of care are compromised. There is no obvious ethical solution to the problem of rationing the medical commons, because any solution involves comparisons between the value of health services provided to different patients with different conditions. None the less, the remedy must lie in some form of collective action: physicians, like the herdsmen in Hardin’s parable, can save the commons only by adhering to a set of mutually acceptable covenants which govern and limit their use of the shared resource.

If health care were “free”, there would be no need to limit its use. Health care is not “free”, because the use of resources (physician time, hospital beds, health care budgets) by some precludes the use of those resources by others. The overall result of failing to adopt covenants that lead to restrained use of resources by well-meaning physicians is unacceptable. If a society mandates universal access to health care, and if all physicians provide their patients with the most beneficial treatments available, then the cost of health care will be unacceptably high to their patients, either as taxpayers, payers of insurance premiums, at the point of care, or in combination. The alternatives are compromises, either with the principle of universal access, or with the principle of unlimited care. Failure to acknowledge these trade-offs can lead to inefficiencies and inequities that compromise both principles: hidden barriers to access such as queues and administrative hassles, exclusion of entire segments of the population, or erecting rigid and arbitrary barriers to introducing new treatments while older treatments of questionable efficacy continue to be used.

This essay begins with a view of this problem from the perspective of society as a whole. Adopting a utilitarian stance, different allocations of resources may be judged by how well they achieve some generally agreed upon measure of social benefit. This representation leads to cost-effectiveness.
analysis as a guide to resource allocation from the societal perspective. The next section describes the fundamental principles of cost-effectiveness analysis as a guide to resource allocation.

The following three sections of the essay contrast the perspectives of patients, consumers (ie, patients ex ante), and physicians on resource allocation. Evidently there are differences among these perspectives regarding the degree to which resource cost should enter into decision making. Concepts that overlap economics and moral philosophy, such as moral hazard and fiduciary relationships, are introduced along the way. I conclude that neither patients nor physicians can be expected to attend to the rationing of the medical “commons” without some form of collective action, but that collective action alone will not succeed without genuine acceptance of the need to ration health care resources by both physicians and patients.

The remainder of the essay concerns the structure of collective solutions, the purpose of which is to permit physicians to fulfill their responsibilities as fiduciary agents to patients, while being responsible gatekeepers of resources. Forms of collective action may range from government regulation to decentralised budgeting of providers, to non-binding clinical guidelines. The promises and limitations of such devices as capitated payments, limited access to technology, and voluntary guidelines, are also considered. An analogy is drawn with public regulation of pollution. The essay concludes with observations regarding the role of cost-effectiveness analysis at the bedside.

The cost-effectiveness paradigm

Suppose that a society wants to provide the maximum aggregate health improvement in its population, but it has limited resources to do so. Each potential health intervention delivered to a defined group of persons with a particular condition yields a health improvement and entails a cost. If health improvement is measured in units which reflect the values of the society, and if costs are measured in units which reflect the extent to which the resource budget is depleted by the intervention, then the maximum societal health improvement can be achieved by applying the following simple decision rule: rank order medical interventions in decreasing order of their expected health improvement per unit expected cost, and adopt them from the top of the rank list to the point on the list that resources are depleted. The ratio of benefit to cost from each intervention represents its “value for money”. To make this decision rule operational even at the societal level, one needs measures of predicted health improvement or benefit, and predicted resource cost. (For present purposes, complexities arising from uncertainty regarding anticipated benefits and costs are set aside, and are summarised in terms of the average, or expected, values of each.)

The health benefit can be measured in units that reflect the preferences of the community, considering their desire for increased longevity but also the value they place on limitation of function, pain, and other dimensions of health-related quality of life. One such measure is the quality adjusted life year (QALY). Quality adjusted life years measure the number of years of remaining life, each adjusted by a preference weight (generally between zero and one, or possibly even less than zero for health states judged worse than dead) that reflects the relative value of the health state on a scale between perfect health (one) and dead (zero). Since the amount of health benefit is uncertain before the intervention, the measure of benefit of a health care intervention can be expressed as an average across similar persons in the target population, in terms of quality adjusted life expectancy. The QALY gain for an intervention should include all health consequences, both positive and negative, and is therefore a measure of net health benefit.

Monetary terms

Costs are usually measured in monetary terms—dollars, pounds, euros—but it should always be remembered that money is only a proxy for the real resources—physicians, nurses, hospital facilities, pharmaceutical development and production—that are consumed in providing the service. From society's perspective, costs are also net of any savings in future health care resources that might have been consumed but for the intervention—such as the costs of treating strokes that are averted because of a blood pressure control programme. From a societal perspective, costs also include resources contributed by patients and family members and other caregivers, including their time, and not just resources financed through health insurance (public or private) or billed to the patient.

In common practice, the ratio reported from cost-effectiveness analyses is the reciprocal of “value for money”, namely the cost per quality adjusted year of life gained. Thus, interventions having low values of this ratio go to the top of the rank list, and those with high values may fall below the line for a particular budget.

Some programmes may actually save more resources than they consume, while still providing a health improvement. Such “cost-saving” programmes automatically go to the top of the rank list because they do not deplete the resource pool; indeed, they add to it. Contrary to popular beliefs, however, very few programmes, even in the area of prevention, are truly cost-saving.

What is the ethical basis for seeking to allocate society’s health care resources with the aim of maximising quality adjusted life expectancy? If each individual measured the value of his or her own health in terms of quality adjusted life expectancy, and if each of them paid for his or her own health care through individual savings accounts, then economic theory would lead them each to allocate their own assets according to the cost-per-QALY rule. Because of the uncertainties about the need for health care during one’s life, however, and for reasons of equity, health care services are typically covered by pooled risk-sharing arrangements, in
the form of insurance or national health care. Under these arrangements, the measure of aggregate QALYs at the societal level entails an interpersonal comparison: the implicit assumption is that a QALY is a QALY, no matter who gets it. Much has been written regarding the ethical justification for such a social value. Citizens behind a primordial veil of ignorance might elect to prespecify that health resource allocations be made so as to maximise their expected utility as viewed from behind the veil. If their preferences after the veil is removed are to maximise their own quality adjusted life expectancies, then a society that maximises aggregate QALYs might appeal to them.

Rawls, however, comes to a rather different conclusion from the same initial premise. He concludes that individuals behind the veil would attach more weight to gains in wellbeing should they emerge with disadvantages, such as poverty or poor health. Thus more weight would be applied to health improvements for those persons in the most poor health. Thus more weight would be applied to gains in wellbeing should they emerge with disadvantages, such as poverty or poor health. Thus more weight would be applied to health improvements for those persons in the most disadvantaged positions in society. Alternatives to QALYs that reflect societal preferences for differential weighting of health gains to the least healthy have been proposed by Nord. Whether a society wishes to count all QALYs equally, or to weight them inversely to the health status of the beneficiary, the result is a utilitarian measure of societal health benefit that could be used to guide resource allocations in health care through an appropriately constructed ratio.

The fundamental question in this essay is this: if there is general consensus that resources should be allocated in such a way as to maximise aggregate health benefit, who is responsible for allocating the resources? Should the patient voluntarily deny himself health services out of a sense of communal obligation to conserve the commons? Should physicians be expected to balance their responsibilities to do the best they can for their patients with a responsibility to be the gatekeeper of the commons at each and every encounter with patients? Or is some form of collective action required, whereby citizens empower their physicians to practise medicine within a system that imposes limits on the resources available to them? My conclusion is that a combination of all three is necessary in order to allocate medical resources efficiently.

The role of consumers (patients ex ante)
As patients, people have a different view of health care costs than they have as consumers. While people may expect that physicians do everything possible for them when they are sick, they complain bitterly when their insurance premiums rise, when the prices of goods and services go up because labour costs to employers reflect rising health care benefits, or when their taxes go up.

Consumers have to pay for their collective use of health care resources, but they don’t want to bear responsibility for the collective costs at the point of their own care. Part of the motivation for individuals is explained by the economic theory of insurance—groups of people can become better off by pooling risks and avoiding major losses in the event they become sick. As a result, citizens demand that health insurance be provided by their employers (or made available at nominal cost) or by government. Moreover, subsidised health insurance enables citizens who would not otherwise be able to afford basic health care to obtain it; there is a redistributive aspect to the provision of health insurance.

When patients have medical insurance, they face different incentives in going to the doctor to seek care. Price becomes less of a factor because the patient does not bear the full cost of care at the time care is sought. In effect, the price facing the patient is lower than the full social cost of health care. The
gap between perceived price and resource cost creates an incentive to utilise more health care than the patient would otherwise be willing to buy. This phenomenon is known as “moral hazard” in the economics literature. It tends to promote excessive use of the “medical commons,” leading to increases in the cost of health insurance beyond what consumers believe is reasonable.

Ironically, the word “moral” in “moral hazard” suggests that patients are at risk of being “immoral” if they overutilise health care services relative to the value of the services they receive. This nomenclature stands in sharp contrast to the predominant ethical value of the services they receive. This nomenclature suggests that patients are at risk of being “immoral” if they overutilise services relative to the maximum they could do. In reality, neither the physician placed in the fiduciary role as agent for the patient, nor the patient facing artificially low prices at the point of care, can be faulted for providing and expecting, respectively, the best health care technology can offer.

The role of physicians
Physicians and patients engage in what economists refer to as principal-agent relationships. This refers to the fiduciary trust that patients (the principals) place in physicians (the agents) to make decisions that maximise the wellbeing of their patients, and to act as advocates for their patients in the health care system. This fiduciary responsibility has been described by Fried as entailing doing whatever is possible to “[preserve] life capacities for the realization of a reasonable, realistic life plan.”

As ideal agents for patients, physicians would consider not only the health consequences of their decisions, but also the economic and psychological consequences for their patients. Out-of-pocket costs, time and inconvenience, and reassurance from diagnostic tests are all part of what patients value. A perfect agent would consider all of these, and weigh them against one another if necessary, as would the patient. A physician who recommends an expensive, new treatment that is not covered by the patient’s insurance but that is equally effective as an insured procedure, would not be acting as a good agent. But suppose the patient preferred a far more expensive, insured, procedure whose only advantage was that it avoided a small scar or a few hours less waiting time. Where does the physician’s obligation as an agent stop? Clearly, physicians often depart from perfect agency even with the most honourable and selfless of intentions.

Principal-agent relationships do not always achieve the goal of perfect proxy decision making and selfless advocacy, if the incentives facing the agent lead to deviations from the decision that is best for the principal. For example, physicians may obtain different levels of remuneration, professional stature, or satisfaction, from their actions, and these incentives compete with the incentives that are aligned with the interests of their patients, such as a sense of obligation and cognisance of external monitoring of their quality of care.

As noted before, physicians are often unable to be perfect agents for their patients because of constraints placed upon them. They have limited time during the day (and even if they work extra hours, their performance may suffer). They are limited by resource constraints in hospitals and laboratories, such as intensive care beds and magnetic resonance equipment.

Sometimes constraints force the physician to make explicit choices between the interests of different patients, as in the setting of emergency triage or in the selection of organ transplant recipients. In these situations, the choices concern patients with names and faces, all of whom are under the care of a single physician or provider organisation. The question at the societal level is whether physicians should be expected to allocate resources between their patients and other nameless, faceless patients who could, perhaps, obtain more benefit if the resources were conserved to benefit them. The question, in other words, is whether physicians should consider the cost-effectiveness of the decisions they make for their patients, recognising that resources are limited. Is it the physician’s responsibility to protect the medical commons?

An entirely different view of the physician’s ethical responsibility would be as an agent for society at large, rather than for individual patients. Under this view of agency, the physician would be compelled to allocate resources in the most cost-effective manner, in order to achieve maximal value for money on behalf of society. In such a world, physicians would make decisions that are in less than the best interests of their individual patients, because not all medically beneficial procedures would be provided. But, on aggregate, consumers would be better off because more health benefit would be achieved.

Or would they? Such an ethic would compromise the fiduciary relationship between patient and physician. It would place the physician in a position of making trade-offs with faceless patients, and then explaining to patients why it was not “cost-effective” to do extra computed tomography (CT) or to prescribe the more expensive drug with a slightly better side effect profile. Physicians trained to do the best for individual patients would balk at this social agency role.

Major sea change
The dilemma, then, is that consumers collectively, concerned about health care costs, have an interest in cost-consciousness in the clinic and at the bedside, while as patients they expect a perfect agency relationship. One possible class of solutions may entail collective action, to constrain (not necessarily regulate!) physicians to act in the collective interest while allowing them to strive to maximise the welfare of their individual patients. This is the type of solution that economists advocate, for example, to induce manufacturers to restrict pollution or to induce consumers to recycle. But in health care, it is not clear that consumers will
accept such collective solutions until they first buy into the premise that resources are limited. I believe that a combination of incentives, constraints, and a major sea change in citizen attitudes toward the finitude of medical care will be required to save the medical commons.

The role of collective action
The goal of a collective solution to protect the medical commons would be served by placing constraints on physicians’ choices such that, even as they strive to do the best they can on behalf of their patients, the result of their doing so leads to cost-effective resource allocation. The constraints would, in effect, force physicians to consider the opportunity cost of their decisions, just as shopping consumers are bound by the cash in their wallet and their credit balance. The consumer seeks to maximise her wellbeing, subject to a budget constraint. The idea is to get the physician to take cognisance of the collective budget constraint.

Many examples of incentives to consider opportunity costs pervade the existing health care systems of the industrialised world. Capitated (per patient) payments to physicians encourage them to use their time efficiently. Capitated budgets for hospitals have a similar effect at the institutional level. One problem with capitation as an incentive at the local or individual physician level, however, is that it mitigates the advantage of insurance and risk pooling. If all providers allocate their separate budgets to maximise health improvement, the overall consequence may be suboptimal if the most cost-effective opportunities to improve health are spread unevenly among providers. Over time, monitoring of caseloads and case-mix could be used to adjust the budget levels and thereby equalise the incremental value of resources in different settings. If physicians consider both QALYs (or another socially desirable measure of health outcome) and cost when deciding how to allocate their budgets, the result would produce the societal allocation of resources that is optimal for all health consumers. In that setting would be led to consider the cost per QALY of alternative decisions, and the result would be the maximum possible production of QALYs by that physician. If physicians consider other values, for example if they give additional priority to the sickest patients as suggested by Nord,11 then the result would reproduce the cost-effective result based on those values.

The current situation in the United States, dominated by managed care, has largely eschewed this sort of decentralised, provider-centred system in favour of direct regulation of medical services. Managed care organisations, bound by bottom-line concerns, keep a close watch on the resource allocations of their affiliated providers. The result has been less this sort of decentralised decision making than a system dominated by prior approval, and sanctions against providers for overutilisation. Whether insurers and managed care organisations actually consider cost-effectiveness in their regulation of care is uncertain.12 If they do, their perspective on both cost and effectiveness departs from that of society at large by virtue of their short time horizons, responsiveness to popular demands over clinical effectiveness, and other factors.13 The result has been a large number of dissatisfied consumers and an even larger proportion of disgruntled physicians.

While some economists might applaud the decentralised, provider-centred solution in theory, it does have a number of practical flaws. For one, physicians would have to fend off patients whose demands for health care services were not being met. There would be a tendency to give more attention to the loudest, most assertive patients, relative to the cost-effectiveness of their claims on the resource pool, and relatively less attention to more passive patients. Perhaps a major barrier to successful implementation of the provider-centred model is the need for buy-in on the part of both physicians and consumers. I will return to this essential ingredient later.

The role of physicians under resource constraints
A responsible physician who cares for a panel of patients, but who is either faced with a resource budget or accountable for the resources he consumes, would be placed in a situation similar to the emergency room physician performing triage. The physician would be responsible for setting or implementing priorities for care and could invoke cost-effectiveness data to help guide these priorities. This would enable physicians to consider the incremental health benefit they could offer to each patient, with perhaps some added moral consideration given to applying resources to the most desperate cases.14 Whatever metric the physicians use, whether implicitly or explicitly, they would be striving to maximise the welfare of their patient population. Like parents taking care of their several children—decisions about what clothing to buy for each, which ones to send to sports camp, which to send to college—physicians could approach these interpersonal comparisons in a caring, compassionate way.

Given the complexity of medical decisions, and in light of the growing recognition of the role of evidence on effectiveness and cost in medical decisions, physicians must work together to develop and interpret the evidence to support a mutually acceptable framework for cost-effective decision making. This entails physician participation at two levels: helping to set the constraints within which they practice, and formulating flexible guidelines to help them allocate the resources under their stewardship. These, then, are the additional roles physicians must play in order to ensure resource allocations that are responsive to patients’ needs and preferences. If physicians accept this role of stewardship for their portion of the medical commons, they can continue to exercise their roles as agents and advocates for patients.
An apt analogy: pollution control

It would be unrealistic, and a violation of the trust patients place in their physicians, to expect physicians to practise cost-effectively without some form of external constraint. The ethical forces acting upon physicians will pull in the opposite direction if physicians are left to make trade-offs between their own patients and other, unseen, patients. The force compelling them to attend to their own patients is, and should be, stronger, with the result that society’s commons problem is left unattended.

The situation is analogous to the control of pollution. As in health care, self-interest (it does not matter whether the self-interest reflects economic or health values) competes with the interests of the greater society. The managers of companies, especially publicly owned ones, have a primary fiduciary obligation to their shareholders. If profit maximisation entails the discharge of pollutants, then it would be a violation of their agency relationship to shareholders to volunteer to regulate pollution more than their shareholders would want. But other members of society, who are not shareholders, have an interest in pollution control. Their interest in the company, while perhaps not zero because the costs of pollution control would be passed on to them in part through higher product prices, would be greatly attenuated by comparison with the adverse effects of the pollution. Should these managers be expected to restrict pollution voluntarily below the levels their shareholders would want? Would it be ethical for them to do so? Recognising that a valid argument could be made that it would be unethical for the managers to fail to abide by these shareholders’ wishes, society adopts pollution control regulations (or pollution discharge quotas or taxes). Responsible companies, while they may lobby for less stringent standards, higher quotas, or lower taxes, generally accept the rationale for collective action to protect the greater society.

The managers of companies, especially publicly owned ones, have a primary fiduciary obligation to their shareholders. If profit maximisation entails the discharge of pollutants, then it would be a violation of their agency relationship to shareholders to volunteer to regulate pollution more than their shareholders would want. But other members of society, who are not shareholders, have an interest in pollution control. Their interest in the company, while perhaps not zero because the costs of pollution control would be passed on to them in part through higher product prices, would be greatly attenuated by comparison with the adverse effects of the pollution. Should these managers be expected to restrict pollution voluntarily below the levels their shareholders would want? Would it be ethical for them to do so? Recognising that a valid argument could be made that it would be unethical for the managers to fail to abide by these shareholders’ wishes, society adopts pollution control regulations (or pollution discharge quotas or taxes). Responsible companies, while they may lobby for less stringent standards, higher quotas, or lower taxes, generally accept the rationale for collective action to protect the greater interest of the community. They then set about their business of maximising profits within the constraints. Sometimes they attempt to bypass the constraints or advocate loosening of the standards, on behalf of their shareholders, but the laws are enforced, and they learn to do the best they can for the shareholders within the constraints.

Concluding remarks: the need for ‘buy-in’

Should physicians do cost-effectiveness analysis at the bedside, even if they are not compelled to ration care by externally imposed constraints? To do so would place them in an untenable position as agents for two, sometimes adversarial, principals—the individual patient in the clinic and the larger community. Patients would lose the undivided advocacy to which they are accustomed, and the pressures on physicians to be responsive would be overwhelming.

Collective action is required to constrain individual providers of health care to practise within their collective means. These constraints could be implemented in and by organisations as diverse as managed care organisations and national health services. Or they could be agreed upon by consensus of provider organisations and citizens. Decision making within these constraints would invite cost-effectiveness analysis by individual providers, but now the choices would be among members of the physician’s own “family” of patients, just as the parent decides how to tend to the needs and wants of her several children. The physician would be free to advocate for each patient, to do the best he can to “[preserve] life capacities for the realization of a reasonable, realistic life plan”.15

Such a means of protecting the “medical commons” could result in resentment on the part of physicians and patients alike, unless both accept the underlying rationale for the constraints. The system of utilisation review by managed care organisations in the United States has spawned a movement in support of a return to physician control of decision making. But a return to unconstrained use of technologies and resources will fail because of the limits at the societal level. The only way out of this dilemma is for citizens and physicians to accept the concept and consequences of resource limits, just as they accept speed limits, zoning laws, and other self-imposed constraints in the interest of the greater good.

Attitude of consumers

How can this acceptance of resource constraints come about? There is some indication that physicians are already willing to accept the idea of limits, as long as they have some degree of control or participation in the process of setting them. Medical specialty societies have begun to incorporate cost-effectiveness into their formulation of clinical guidelines. A recent study showed good concordance between the rankings by a consensus panel of physicians of appropriateness of a medical technology in different indications—coronary angiography after myocardial infarction—and cost-effectiveness ratios for these indications.18

The biggest obstacle to cost-effective resource allocation along these lines is the attitude of consumers. People who expect everyone to have access to all possible medical care regardless of cost are bound to be disappointed. In the United States, the sustenance of this myth has come at a high price. Even in times of economic prosperity, an increasing proportion of Americans is uninsured, and millions of them have limited or no access to medical care. Some of the most cost-effective but least glamorous medical interventions, such as screening for colorectal cancer and adult vaccinations against influenza and pneumonia, are underutilised in favour of the procedures consumers want and demand. An informed populace, aware of the finite benefits of health care services and of the rationale for constrained choice in the clinic and at the bedside, is an essential ingredient if physicians are to be burdened with the responsibility for cost-effective decision making. This evidence suggests that the medical profession will accept that responsibility if it is given the authority to set the rules by which its
members are to play, and if their patients are at peace with the principle of living within their means.

Milton C. Weinstein, PhD, is Director of the Program on the Economic Evaluation of Medical Technology, Center for Risk Analysis, School of Public Health, Harvard University, Boston, Massachusetts, USA.

References

News and notes
Life without Disease: Advanced European Bioethics Course

This year’s Advanced European Bioethics Course, Life without Disease, will be held from 15 - 17 November in Nijmegen, the Netherlands. Subjects will include: stem cell research; xenotransplantation; tissue engineering; cloning; reproductive planning, and pharmacogenetics.

For more information contact: N Steinkamp, Catholic University Nijmegen, 232 Department of Ethics, Philosophy and History of Medicine, PO Box 9101, 6500 HB Nijmegen, the Netherlands. Tel: int. 31-24-3615320; fax: int. 31-24-3540254; email: n.steinkamp@efg.kun.nl

www.jmedethics.com