into something else. If they are not for profit they'll hang in there, for ever stretching out their reserves and taking losses until they are driven into bankruptcy. We have had some hospital mergers—a very good thing, I think—closing a hospital by redirecting its activities to less costly uses such as stroke rehabilitation, doctors' rooms, and low tech care while transferring all acute care.

A health maintenance organisation providing the total range of care can reallocate resources within its organisation from the inpatient sector to the outpatient system

We have way too many specialists. Health maintenance organisations contract with a limited number of specialists to ensure that they have enough patients to be fully proficient in the specialty they trained in. One of the keys to a successful health maintenance organisation is to be able to adjust the specialty mix to the needs of the population served. In the United States we are going to see more highly trained specialists unable to get a job.

I think medical academia is very slow to get the word and slow to adapt and to adjust. Two to three years ago few graduates went into primary care. Now we are not short of primary care physicians. We have the same ratio of primary care physicians to the population as in the United Kingdom—roughly 1800 people per doctor. We would not have been able to maintain that if the previous trend had continued, but now medical students are coming to realise that what's needed is primary care physicians. If they go into radiology or cardiology they might not be able to get a job.

PN: A lot of what you have discussed sounds very familiar and is also happening in Britain. For example, general practitioner fundholders are forming consortiums or "multifunds" and purchasing the full range of care in the total purchasing projects. There is an emphasis on a shift of activity and resources from secondary care into primary care. Some hospitals have adopted critical paths and use clinical guidelines. Some purchasers are moving to preferred provider relationships, and in some major cities—notably London—there are proposed hospital mergers. Are we heading in the same direction as the United States?

AE: I'd like to think that there is some convergence. I think there is a long way to go, as you are coming from a very socialised and monolithic kind of system in which the element of consumer choice has not been very strong. My impression is that it is not a part of British culture to want to change doctor or the services very much. You take the general practitioner who was assigned to you. I don't know how much a cultural change could occur in Britain. It would be a big change for people to move from one general practice to another because they felt the service wasn't good.

In both cases market forces can be used to provide incentives for doctors and other medical personnel to improve the quality of care and reduce the cost and therefore give better care. I had a call from a journalist the other day from Britain who said: "This is terrible. We have cut throat competition going on." But from the patients' point of view that's wonderful. She said: "the prices are coming down"—that's marvellous because you have limited resources and more people can receive better care. Our experience is showing there is lots of room for innovation in medical care processes.

I thank Professor Enthoven for the interview; the Faculty of Public Health Medicine, Royal College of Physicians, King's Fund College, and the North West Thames Regional Health Authority for travel fellowships; and Sheila Adam and Jennifer Dixon for encouragement.

(Accepted 4 April 1995)

Dilemmas in rationing health care services: the case for implicit rationing

David Mechanic

With tension between the demand for health services and the cost of providing them, rationing is increasingly evident in all medical systems. Until recently, rationing was primarily through the ability to pay or achieved implicitly by doctors working within fixed budgets. Such forms of rationing are commonly alleged to be inequitable and inefficient and explicit rationing is advocated as more appropriate. Utilisation management in the United States and quasi-markets separating purchasing from provision in the United Kingdom are seen as ways of using resources more efficiently and are increasingly explicit. There is also advocacy to ration explicitly at the point of service. Mechanic reviews the implications of these developments and explains why explicit approaches are likely to focus conflict and dissatisfaction and be politically unstable. Explicit rationing is unlikely to be as equitable as its proponents argue and is likely to make dissatisfaction and perceived deprivation more salient. Despite its limitations, implicit rationing at the point of service is more sensitive to the complexity of medical decisions and the needs and personal and cultural preferences of patients. All systems use a mix of rationing devices, but the clinical allocation of services should substantially depend on the discretion of professionals informed by practice guidelines, outcomes research, and other informational aids.

King's Fund Institute, London W2 4HT
David Mechanic, *visiting fellow*

BMJ 1995;310:1655-9

Medical care has always been rationed by the supply available, by its distribution, and by the public's ability to pay. As medical care has become more important in people's lives, and as its capacity to impact on health has grown, government has taken increasing responsibility either for providing medical services directly or for mandating them through an insurance system. Governments in all nations seek means to limit public expenditure and mandates for health services. Explicit approaches include fixed global budgets and limits on the benefit package and eligible providers. Rationing also occurs implicitly through cost sharing, waiting lists, and requiring professionals to work within a constrained budget.

Rationing is a "complex interaction of multiple

Summary points

- Medical care has always been rationed, but rationing strategies are changing
- All systems of care use mixed rationing approaches; the key issue is finding a proper balance
- Advocacy to allocate care on the basis of more explicit methods such as public preferences, clinical guidelines, and outcomes research is increasingly common
- Explicit rules are unresponsive to rapidly changing medical knowledge and variations in patients' preferences and tastes
- Explicit rationing inevitably gives preferences to some who care less about treatment than others who are excluded, contributing to social conflict
- Trust between doctor and patient holds the system intact; incentives that violate trust should be avoided
- Social judgments are readily confused with subjective judgments of medical necessity, and it is important to guard against unconscious preferences reflecting class, sex, and other social biases
- Though explicit controls are needed over financing and diffusion of unproved technologies, implicit rationing at the patient level offers the most sensitive way of responding to differences among patients in their needs and preferences

decisions, taken at various levels."[1] All systems of care use mixed rationing approaches[2] but the relative balance is a matter of continuing debate. Increasingly, implicit rationing has been under attack as uninformed, arbitrary, and inequitable. Instead, it is argued that explicit strategies such as contracting and purchasing arrangements, rating systems that establish people's preferences and the value they place on varying medical outcomes, determinations based on quality adjusted life years, outcomes research, and practice guidelines should dictate allocation decisions. These are useful aids but not useful directives. I maintain, in contrast, that though some explicit controls are needed over financing and the diffusion of expensive new technologies, explicit rationing at the micro level will increase tensions, conflict, and instability. I begin by illustrating some of the difficulties in the managed care sector in the United States.

Rationing through managed care

The idea of patients freely pursuing an insurance entitlement is undergoing substantial change in the United States under managed care. Managed care restricts the entitlement of covered services to those that are deemed medically necessary by managed care administrators. This emerging concept modifies "medical necessity" from solely a professional determination to an administrative judgment. Unlike in the British NHS, procedures that are part of the insurance contract cannot be withheld simply because of cost or limited resources, but the concept of medical necessity is vague enough to incorporate economic considerations.

Managed care now exists in many forms, representing different rationing approaches.[3] In its traditional form in prepaid group practice and other types of health maintenance organisations the restraining mechanism is capitation and the need to stay within established budgets. This results in a type of implicit rationing traditionally seen in the NHS in which each clinician makes judgments, aware that resources are limited. But only a minority of patients in the United States are covered by capitation and the implicit allocative processes that follow from it.

UTILISATION MANAGEMENT

Most Americans, however, are affected by utilisation management, which includes precertification for admission to hospital, concurrent review of inpatient length of stay, case management of high cost cases, and second surgical opinions. These devices are joined in various ways and often combined with capitated contracts and employee assistance programmes. Depending on how managed care is administered, it commonly constitutes a form of implicit rationing in that allocations continue to depend on discretionary judgments of physician reviewers. Alternatively, to the extent that utilisation managers work on the basis of protocols executed by non-medical personnel, rationing shifts to a more explicit form. Two features of utilisation management deserve special attention. Firstly, it shifts authority from the practising clinician to others. Secondly, it substitutes a more formalised and explicit determination for traditional clinical implicit decision making.

Though managed care companies depend on medical authority in establishing criteria, administrative decisions can overrule clinical judgments. All reputable utilisation management companies use physicians to establish standards of care and help in decision making, and doctors typically become concerned in some way in initial reviews or appeals. In one sense, then, medical authority over medical work remains. However, doctors now acting in administrative roles, and often as agents of profit making companies, have the power to countermand the judgment and wishes of the patient's doctor and deny access to services.

Most companies are pragmatic and tread carefully to avoid alienating doctors. They seek to achieve their goals more by deterrence and negotiation than by the exercise of power, and unqualified denial is fairly infrequent. But it is not unusual for utilisation reviewers to refuse services that the clinician believes essential, most commonly a longer inpatient stay. The conflict in competing definitions of medical necessity becomes something of a "tug of war," as little evidence typically exists to back either position definitively. There are appeals procedures and alternative ways clinicians can advocate for their points of view, but the process of utilisation management, and advocacy and appeals processes even more, imposes considerable costs on the practising physician in time, frustration, and loss of professional autonomy.

ADVOCACY AND APPEALS

We know little about how doctors deal with advocacy and appeals opportunities, but the process is likely to be particularly burdensome for single handed or small practice units lacking the staff to monitor and follow up these matters. A doctor without such staff may have to personally make several attempts to get the utilisation management company on the phone, be put on hold and lose time waiting, and may need to make repeated contacts. Such doctors might be reluctant to make the necessary advocacy efforts. Though quite different from the situation in Britain, comparable issues pertain to general practice fundholders and how they perform their new responsibilities.

When managed care organisations refuse to authorise a service it remains unclear what obligations doctors have to patients to inform them of disagreements about their care. Informing patients protects doctors against

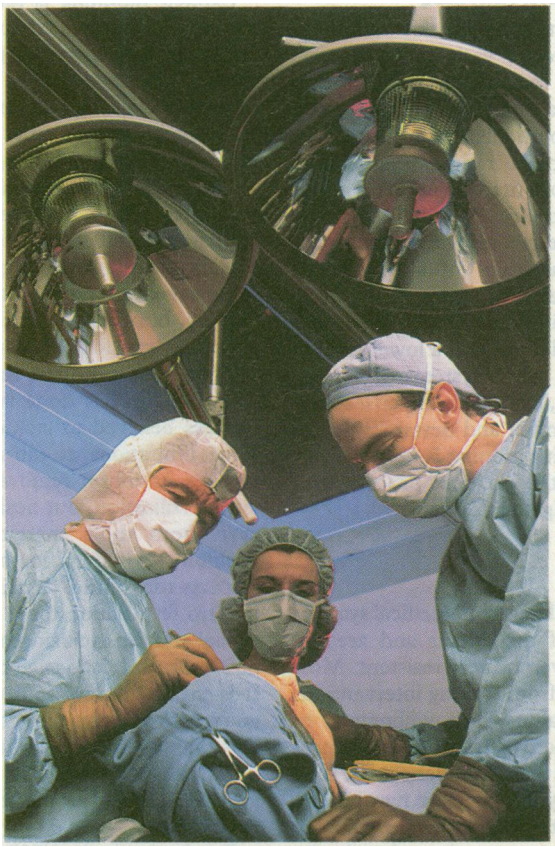
possible malpractice litigation should negative outcomes result and may help resolve clinicians' guilt that they are not doing all they think they should for their patients' welfare. It also provides patients with an opportunity to influence the managed care company directly or through their employer, or to engage an attorney should they wish. Revealing such conflicts about what is medically necessary is uncomfortable, however, and may lead the patient to question the doctor's authority or effectiveness.

It has been suggested that if patients are informed that care will be rationed as part of their insurance contract doctors could be absolved of legal responsibility to inform them about decisions to withhold care.[4] The difficulty is that patients pay little attention to the formal language of insurance contracts and access issues typically are not salient until people face serious illnesses and are denied the care they believe they need. Thus it seems appropriate to expect doctors to inform patients of realistic options and when services they believe would be helpful are being withheld because of managed care decisions.[5]

Utilisation management, in adding a layer of management over direct doctor-patient transactions, moves towards explicit rationing at the point of service. Though theoretically utilisation management seeks to dispense only with unnecessary care, the concept of necessity is vague and practice variations are large enough to leave much to argue about.[6] As managed care companies tighten up on costs, increasing conflict and controversy are likely. Differences of opinion between clinicians and utilisation reviewers will contribute to patient dissatisfaction, hostility, and lack of trust.

Strengths and weaknesses of implicit rationing

In systems where services are provided directly, as in the NHS, patients theoretically have access to all possible medical interventions. But unlike in the United States, where patients have a legal basis for demanding a needed procedure covered under their insurance policy, the NHS consents to provide only those medically necessary services that can be managed within the resources available. Patients who need other than an emergency service can join a waiting list, but they have no absolute right to receive the service in a timely way. The waiting list serves to fit acuteness of need to the resources available.[7] Critics, noting large variations in waiting lists and access to services in varying health districts, argue for making allocation of resources more explicit.



MORGAN/SPL

Initial restrictions by health districts in relation to some cosmetic operations have been relaxed

CRITICISMS OF IMPLICIT RATIONING

A common criticism of implicit rationing is that knowledgeable, sophisticated, and aggressive patients are more able to have their needs satisfied than docile patients. Implicit rationing is also seen to suffer from the discretion it gives doctors who may act on personal preferences or ignorance of medical advances. Social judgments are readily confused with subjective judgments of medical necessity and preferences creep in unconsciously, reflecting class, sex, and other social biases. These problems require attention and protective measures but without impairing the ability to make needed discretionary judgments. In an admittedly extreme case Halper quoted a respected British nephrologist who saw gainful employment as an appropriate criterion for haemodialysis because "only the minority wish to live on charity."[8] More commonly, doctors make assumptions about benefit based on judgments about intelligence, family circumstances, personality traits, and the like. Such judgments may in many instances correctly predict future outcomes but commonly reflect social biases more than empirical reality.

Arguments for and against explicit rationing

Arguably more definitive rules are needed to ensure the wise use of resources. For example, Sheldon and Maynard maintain that, "If we want a service that uses the public's money to promote health in an efficient and equitable way... it is important to get involved in rationing to insure that it occurs in a responsible and just fashion rather than the current process, which is largely uncharted and the product of clinical discretion which creates major variations in practice and patient access."[9]

Explicit guidelines, however, are likely to fall short relative to the complexity of circumstances surrounding serious illness and comorbidities or to be so complex and detailed that they are impracticable. Well developed guidelines, however, focused on high cost areas can be invaluable aids in the responsible exercise of discretion—for example, when establishing priorities for revascularisation[10,11] or use of intensive care beds.[12]

Building a new culture of medical practice

A culture of medical practice is needed that is accountable and takes responsibility to use resources wisely and consistent with unfolding knowledge of best practice and cost-benefit outcomes. This is a long range process and requires involvement of medical schools, postgraduate education, and emerging peer review structures in both hospital and primary care. Impatience and the temptation to intervene aggressively in a regulatory way are understandable but probably also counterproductive. Practice guidelines seen as important educational and practice aids will be incorporated more readily into clinical thinking than if imposed externally by regulatory authorities as a strategy to control medical decision making.

THE CASE OF RENAL DIALYSIS

End stage renal disease offers an atypical but instructive example. The British initially judged implicitly that younger people should have preference in access to haemodialysis, and many fewer elderly patients received this service than in other countries.[13] But the understanding was not absolute, and patients of all ages could receive chronic dialysis. Over time the proportion of elderly patients who were dialysed substantially increased. Such discretion may be viewed as inequitable. But a flexible standard allows taking into account the extraordinary variation in aging among people of the same chronological age, changing opportunities for successful intervention, differences in preferences, and a wide variety of other contingencies. The seemingly discriminatory outcomes may in fact be more equitable than formally applying explicit rules.

The assumption that explicit rules are inviolable is also unrealistic. In 1961, while teaching at the University of Washington Medical School, I was lunching with the chief nephrologist, a pioneer in establishing the first chronic dialysis unit for patients with end stage renal disease. The unit had developed an evaluation process in which community representatives participated to allocate the scarce places available on the artificial kidney. There was much criticism of the procedures used, but it was the process then in place. During lunch the chief was informed that the state's senior senator, a powerful figure in Washington and a strong friend and promoter of the medical school, was insistent that a friend with kidney failure should have dialysis, though it was unclear that he could receive this service through the usual process. While we sat there a plan was made to bypass the allocation committee by giving this patient dialysis as a research subject. I suspect that the rich and powerful if sufficiently motivated will always find ways to circumvent explicit criteria.

Destabilising effects of explicit rationing

Problems with implicit rationing are not trivial and measures are needed to reduce variabilities. Nevertheless, implicit rationing reduces tensions arising from scarcity by taking into account the determination of people to receive a particular procedure. An important weakness of explicit rationing is that it inevitably gives preference to some who care less about treatment than others who are excluded. Thus it results in many disaffected people who are a continuing force challenging either the rules of allocation or decisions to withhold greater investment in the area. Implicit rationing, despite its imperfections, is more conducive to stable social relations and a lower level of conflict. It is doubtful that tough systems of explicit rationing can be maintained, except during crises such as war, without focusing conflict and destabilising the medical care system. Explicit rationing is also likely to confront government and the political process with unrelenting agitation for budget increases.

Role of physician trust

Patients even in the most "liberated" medical systems still accept the authority of the doctor and are not inclined to make a fuss or even be very insistent. Most patients can be discouraged from seeking interventions that have only marginal value. Such willingness to accept medical judgments derives from the trust most patients have in doctors and their confidence that their doctors would not knowingly take steps to harm them. In all countries, however, there are people who are distrustful, assertive, and less accepting of authority and the trend is probably in that direction. Such patients when insistent are probably more likely to receive the interventions they seek because doctors typically are uncomfortable with the tensions these patients introduce. This is particularly so when the intervention at issue is seen as efficacious but there is just not enough to go around.

Role of patients and their families

Most discussions of rationing proceed as if the patient is passive and as if his or her preferences, tastes, and demands are of little importance in the allocation process. Patients vary a great deal in their illness behaviours, the extent to which they inform themselves about their illnesses and possible options, and the roles they take in their own care. Some patients are passive and content to put themselves fully in the care of their doctors. Others read medical publications, talk to other patients, see different doctors for second and third opinions, and take considerable control over their own treatment plans. Patients also vary greatly in their determination to overcome the adversities of their illnesses, cope as normally as possible, and continue with their lives. People have very different preferences for added increments of life itself.

Insensitivity of explicit rules

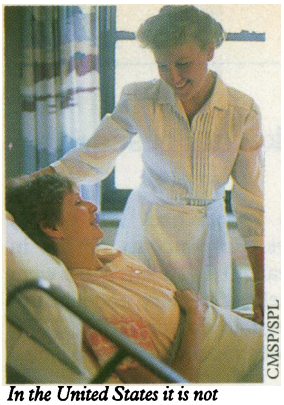
It is no easy task sensitively to take these aspects into consideration when allocating scarce life enhancing or lifesaving resources, but it is not unreasonable that they should be and it cannot be done in advance. Implicit rationing provides the flexibility to do so whereas explicit mandates rigidify alternatives through superficial assessments of equity-efficiency trade offs. Explicit rules inevitably will be insensitive to the innumerable differences among people and circumstances.

Instability of explicit rationing approaches

Explicit rationing is inevitably unstable because of the ability of small groups to evoke public sympathy and support in contesting government decision making. Those who care deeply but are denied access will inevitably challenge the explicit judgment through the mass media and in other ways, undermining support for purchasing decisions and pushing the health system towards more flexible implicit approaches. This is already evident in the case of the reforms in the United Kingdom despite the rhetoric about the need for more explicit decision making. New arrangements allow some negotiation on volume, waiting time, and related matters, but there is little evidence that the allocation process is transformed in any fundamental way.

There was a notion that health districts might do what Oregon has done and rule out the provision of services believed to lack efficacy. Initially, there were some limitations at the margins in relation to such services as tattoo removal, cosmetic surgery, and in vitro fertilisation, but health districts have pulled back even from these marginal restrictions as absolute prohibitions.[14] Even in those few cases where health districts took initial decisions to exclude, for example, in vitro fertilisation these measures have been controversial and the exclusions relaxed. Such examples show how difficult it is to exclude even marginal procedures. It is easier and less controversial to restrain resources going into any area and control access through waiting lists than to impose explicit prohibitions. Waiting lists do not avoid controversy, particularly as waiting times grow long, but they are less contentious than absolute prohibitions.

Purchasing in the NHS is potentially a form of quasi-explicit rationing. It is explicit in the sense that any


CMSP/SPL

In the United States it is not unusual for utilisation reviews to refuse services—most commonly a longer inpatient stay

purchasing authority may choose not to buy a particular form of service or, alternatively, to buy a service in such restricted form that it is unavailable to most people in need. But even in such restricted cases clinicians pick and choose who gets priority and how long patients must wait. Block purchasing is likely to continue as the dominant form, depending on the professionalism and good will of clinicians to use available resources efficiently and fairly. Medical discretion remains strong and largely undisturbed as long as doctors agree to be allocators as well as advocates.[15] Theoretically, anything is possible if medical authority sanctions it. Neither central government nor health districts are likely to have the stomach to ration explicitly in substantial ways and thus they remain dependent on trust that physicians will use discretion in a politically acceptable and hopefully wise way.

The erosion of trust

Implicit rationing works because patients trust that doctors are their agents and have their interests at heart. Trust varies from one cultural setting to another and seems stronger in the United Kingdom than in the United States, but all medical systems depend substantially on it. Trust holds the system together in the face of economic and other tensions, and in its absence mechanisms of needed control are expensive, burdensome, and uncertain. Trust in an important sense is a substitute for a cumbersome regulatory bureaucracy.

With increased attention of the media to conflicting medical views, higher level of education and sophistication of patients, and erosion of respect for authority trust has become more fragile. Patients are more commonly aware of options their doctors do not suggest, and—though most are still docile—patients are increasingly challenging medical judgment. The easy availability of medical information makes it likely that some patients will be more acquainted with new developments than their doctors and may be less accepting of the idea that little more can be done. Maintenance of trust and the authority of the doctor are thus likely to depend on patients' perceptions that their concerns are primary in the doctor's attitudes and actions.

New organisational arrangements are increasingly challenging trust in the American context, and present trends suggest similar trends are likely in the United Kingdom as well. The emergence of a vigorous profit oriented sector in health care results in pressures to reduce costs and maximise revenues. Considerable attention has been focused on doctors' investment in medical facilities to which they refer patients, but much less public attention has been given to the remuneration incentives for doctors to stay within utilisation limits. In many for profit health maintenance organisations a large proportion of the doctor's earnings depends on meeting expected cost targets.[16] Most patients are uncomfortable with the idea that their doctor must balance their needs against the needs of others.[17] If patients truly knew the extent of developing conflicts of interest built into existing financial and organisational arrangements[18] their trust would be much diminished. Yet it is inevitable that patient awareness on these matters will grow.

Conclusion

Open decision making and equitable distribution in relation to need are important and widely held values. But they are difficult to implement in nations with population heterogeneity and an increasingly strong rights culture associated with unwillingness to bear new tax burdens. Dealing with differences in culture, tastes, and wants requires discretion and flexibility. Differences made too explicit are likely to lead to resentments and conflicts. Thus, though it is possible to have an informal understanding that the same vigour of intervention for the sick old and younger patients is inappropriate, making the policy explicit and applying it uniformly will inevitably result in acrimony difficult to manage politically. Wise bureaucrats have always understood that organisations can easily be disrupted by rigid adherence to rules. A certain flexibility is needed for any organisation to thrive.

In short, rationing at the micro level must be left for doctors and patients to work out among themselves. Informal resolutions must take place within explicit constraints but once the boundaries are set more is gained by muddling through than by trying to establish all the rules beforehand. Seriously ill patients pose substantial complexities and, depending on how illness, culture, and personality combine, may require different care. Herbert Simon, one of the few people to receive the Nobel prize for organisational wisdom, noted years ago that managers must satisfice because they lack the wit to optimise.[19] Interest in making rationing explicit arises from the illusion that optimisation is possible. Implicit rationing embedded in an appropriate value framework offers the best among admittedly imperfect alternatives.

I thank Nick Mays and Julian Le Grand for helpful comments.

1 Klein R. Dimensions of rationing: who should do what? *BMJ* 1993;**307**: 309-11.
2 Mechanic D. Approaches to controlling the costs of medical care: short-range and long-range alternatives. *N Engl J Med* 1978;**298**:249-54.
3 Mechanic D. Managed care: rhetoric and realities. *Inquiry* 1994;**31**:124-8.
4 Hall MA. Informal consent to rationing decisions. *Milbank Q* 1993;**71**:645-68.
5 Mechanic D. Trust and informed consent to rationing. *Milbank Q* 1994;**72**: 217-23.
6 Mechanic D. Professional judgment and the rationing of medical care. *University of Pennsylvania Law Review* 1992;**140**:1713-54.
7 Frankel S, West R. *Rationing and rationality in the National Health Service: the persistence of waiting lists.* Houndsmills: MacMillan, 1993.
8 Halper T. *The misfortunes of others: end-stage renal disease in the United Kingdom.* Cambridge: Cambridge University Press, 1989:114.
9 Sheldon TA, Maynard A. Is rationing inevitable? In: *Rationing in action.* London: BMJ Publishing Group, 1993:3-14.
10 Naylor CD, Basinski A, Baigrie RS, Goldman BS, Lomas J. Revascularisation panel, consensus methods group: assessment of priority for coronary revascularisation procedures. *Lancet* 1990;**335**:1070-3.
11 Naylor CD, Levinton CM, Wheeler S, Hunter L. Queueing for coronary surgery during severe supply-demand mismatch in a Canadian referral centre: a case study of implicit rationing. *Soc Sci Med* 1993;**37**:61-7.
12 Knaus WA, Wagner DP, Lynn J. Short-term mortality predictions for critically ill hospitalized adults: science and ethics. *Science* 1991;**254**:389-94.
13 Aaron HJ, Schwartz WB. *The painful prescription: rationing hospital care.* Washington, DC: Brookings, 1984.
14 Redmayne R, Klein R, Day P. *Sharing our resources: purchasing and priority setting in the NHS.* Birmingham: National Association of Health Authorities and Trusts, 1993. (No 11.)
15 Mechanic D. *From advocacy to allocation: the evolving American health care system.* New York: Free Press, 1986.
16 Hillman AL. Financial incentives for physicians in HMOs: is there a conflict of interest? *N Engl J Med* 1987;**317**:1743-8.
17 Mechanic D, Ettel T, Davis D. Choosing among health insurance options: a study of new employees. *Inquiry* 1990;**27**:14-23.
18 Rodwin M. *Medicine, money and morals: physicians' conflicts of interest.* New York: Oxford University Press, 1993.
19 Simon H. *Administrative behavior.* 2nd ed. New York: Free Press, 1957.

(Accepted 21 March 1995)