Principles of justice in health care rationing

Richard Cookson and Paul Dolan University of East Anglia and University of Sheffield, respectively

Abstract
This paper compares and contrasts three different substantive (as opposed to procedural) principles of justice for making health care priority-setting or “rationing” decisions: need principles, maximising principles and egalitarian principles. The principles are compared by tracing out their implications for a hypothetical rationing decision involving four identified patients. This decision has been the subject of an empirical study of public opinion based on small-group discussions, which found that the public seem to support a pluralistic combination of all three kinds of rationing principle. In conclusion, it is suggested that there is room for further work by philosophers and others on the development of a coherent and pluralistic theory of health care rationing which accords with public opinions.

(Keywords: Health care; rationing; medical ethics; justice; need)

Introduction
Textbooks and handbooks of medical ethics typically recommend that medico-moral decisions should be guided by four basic philosophical principles: (i) respect for autonomy, (ii) beneficence (“the patient’s interests come first”), (iii) non-maleficence (“above all do no harm”), and (iv) justice. This paper is about the fourth of these principles: in the context of day-to-day health care priority-setting or rationing decisions by clinicians and administrators, what does “justice” really mean?

Three different principles of justice in health care rationing decisions are commonly discussed in the academic literature: need principles, maximising principles and egalitarian principles. This paper compares and contrasts these three principles with reference to a hypothetical rationing dilemma involving four identified patients. One purpose of this exercise is to tease out the similarities and differences between each principle. A second purpose is to enable comparison between these theoretical principles and the views of the UK general public, whose opinions about this hypothetical example have previously been investigated in an empirical study based on small-group discussions.

This paper focuses on principles of substantive justice in making decisions about who should get what health care and when, rather than principles of procedural justice about what decision making process should be followed. This means, for example, that contractarian principles of distribution according to a process of voluntary transactions are not discussed. In addition, there may be other less well-known substantive principles of justice not discussed in this paper. Given time and resource constraints, it was not possible to perform a systematic and comprehensive review of all principles of justice that have ever been proposed, since bibliographic information on the relevant humanities and social sciences literature is widely dispersed among many different bibliographic sources.

The paper does not address broader questions about health care rationing, such as (1) whether or not rationing is inevitable, (2) whether or not rationing should be explicit, and (3) whether or not different rationing principles should apply in different contexts (for example, “macro” spending choices versus “micro” choices between patients). Nor does it address the question of whether or not it is possible and/or desirable for a society to agree on a single set of principles to guide health care rationing decisions in different contexts. Some academics argue that it is naive to expect that rationing principles will be consistently followed in practice, and that the best we can hope for is to identify better procedures for making decisions (for instance involving more public participation). This position has been referred to as “muddling through elegantly”.

Even if this position is accepted, however, there will still be a role for philosophical discussion of substantive rationing principles, to help clarify the thinking of those who participate in the process of rationing.

The rationing principles
Table 1 lists the main substantive rationing principles that we identified from the literature, classified into three major categories: need principles, maximising principles and egalitarian principles. Need principles require that health care be distributed in proportion to “need” (for example, in terms of immediate ill health). Maximising principles require that health care be distributed so as to achieve maximum “benefit” (for example, in terms of population health). Finally, egalitarian principles require that health care be distributed so as to reduce “inequality” (for example, in terms of lifetime health).

Commonly discussed variants in each category are also listed, with an associated implication for the hypothetical rationing choice between four
In our rationing exercise, this narrow definition of "need" leads to quite different substantive rationing principles. The most common strategy is to define need in terms of the degree of ill health. For example, it can be argued that an immediate threat to life (for example, within the next year or so) is the most urgent and pressing degree of ill health. For example, it can be argued that many rationing decisions will require a binary judgment about whether or not to rescue those in immediate need. As in our hypothetical example, it may not be possible to rescue all those in need, and rescuing one person may have the consequence that other people cannot be rescued. When considering need principles, however, it is important to emphasise the point that many rationing decisions will require a comparative judgment about the relative degree of need, as well as a binary judgment about whether or not a need exists.

A third (rarely discussed) possibility is to broaden even further the interpretation of need as ill health by taking a broader time horizon—for instance by looking at the individual’s whole lifetime of ill health rather than just his or her immediate situation (for example, ill health in the next year or so). If we did this, then Marinder would fall down the ranking relative to Steve and Joanne. This is because she has already enjoyed a relatively long and healthy life and so appears relatively less needy from a lifetime perspective than from an immediate perspective.

### Table 1: Substantive principles of justice in health care rationing

<table>
<thead>
<tr>
<th>Class of principle</th>
<th>Variants that have been proposed in the literature</th>
<th>Recommendation in our exercise*</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Need principles</td>
<td>Distribute in proportion to degree of immediate threat to life</td>
<td>Daniel, Joanne</td>
</tr>
<tr>
<td></td>
<td>Distribute in proportion to degree of lifetime ill-health</td>
<td>Daniel, Joanne, Steve, Marinder</td>
</tr>
<tr>
<td>B. Maximising principles</td>
<td>Maximise health</td>
<td>Daniel, Marinder &amp; Steve equal, Joanne</td>
</tr>
<tr>
<td>C. Egalitarian principles</td>
<td>Equalise lifetime health expectancy (&quot;fair innings&quot; argument)</td>
<td>Daniel, Joanne, Steve, Marinder</td>
</tr>
<tr>
<td>D. Combination principles</td>
<td>Combine maximise health with equalise lifetime health expectancy</td>
<td>Daniel, Joanne, Steve, Marinder</td>
</tr>
<tr>
<td></td>
<td>Combine a need principle with a maximising principle</td>
<td>Daniel, Joanne, Steve, Marinder</td>
</tr>
</tbody>
</table>

*This is based on various working assumptions described in the main text of this paper, in particular (1) that Steve and Marinder have an equal degree of immediate ill-health, (2) that Steve will gain the most health from treatment (over his entire lifetime), that Joanne will gain the least, and that Daniel and Marinder will gain the same amount, (3) that treating Marinder would have a substantial indirect health benefit for others by freeing up long term care resources; and (4) that Joanne had more choice about her health predicament than Steve.

**NEED PRINCIPLES I: NEED AS ILL HEALTH**

Distribution of health care according to need is perhaps the most widely discussed rationing principle in both academic and non-academic debates. It is especially popular among clinicians, who see themselves as the expert judges of need and hence typically use the phrase “clinical need”. The British Medical Association guidelines on medical ethics, for example, suggest that most clinicians accept the principle of distribution according to clinical need. And the Chairman of The Royal College of General Practitioners, Sir John Toby, has stated that: “We don’t believe there should be discrimination on any grounds other than on clinical need”. However, unless the concept of “clinical need” is clearly defined using substantive criteria, this principle reduces to the rather unpalatable procedural principle that any rationing decision must be correct so long as a clinician has taken it.

Different definitions of “need” lead to quite different substantive rationing principles. The most common strategy is to define need in terms of the degree of ill health. For example, it can be argued that an immediate threat to life (for example, within the next year or so) is the most urgent and pressing form of ill health, and that saving (or prolonging) life should almost always take priority over enhancing life. In our rationing exercise, this narrow definition of need as immediate threat to life would imply that Daniel should have priority, on the grounds that Steve, Marinder and (arguably) Joanne are not facing an immediate threat to life. Joanne would presumably then take second place, since she faces the next most immediate threat to life as she starts to develop the AIDS virus.

A second, broader definition of need as ill health would encompass immediate pain and suffering (for example, Marinder’s bad hip) as well as immediate threat to life. Need can then be interpreted as the individual’s immediate degree of ill health. On the basis that Joanne is not immediately suffering, since she has only just been diagnosed HIV positive, it seems plausible that Daniel is the most immediately ill (and hence should have priority according to this second need principle), followed by Marinder, then Steve, then Joanne.

The two need principles discussed above are both sometimes called the “Rule of Rescue”. The idea behind this phrase is that society has a duty to do everything possible to rescue all those facing immediate threats to life and/or health. This phrase can be misleading, however, since it suggests that health care rationing is a binary question about whether or not to rescue those in immediate need. As in our hypothetical example, it may not be possible to rescue all those in need, and rescuing one person may have the consequence that other people cannot be rescued. When considering need principles, however, it is important to emphasise the point that many rationing decisions will require a comparative judgment about the relative degree of need, as well as a binary judgment about whether or not a need exists.

www.jmedethics.com
Figure 1: The rationing exercise

Patient S
18 year old Steve is the victim of a car accident. He has severe facial scarring and psychological problems as a result. Plastic surgery would correct the scarring.

Patient J
Joanne is 42 years old and has no dependents. She has just been diagnosed as HIV positive as a result of her drug taking. She no longer takes drugs. Drug treatment is available which is 75% effective. This could extend her life expectancy and minimise symptoms.

Patient M
Manser is 65 years old and has been waiting 16 months for a hip replacement. Soon, she will be no longer able to live alone. Her only son lives 200 miles away. The hip replacement would allow her to live independently.

Patient D
18 year old Daniel has cancer. He has a 50/50 chance of survival. There is however a new drug treatment available that has been partially tested on a limited number of cases.
It is also possible to define need in terms of the potential to benefit from health care, as opposed to the initial level of ill health. However, we will be in a better position to discuss need definitions of this kind later on, once we have looked at maximising principles which are more directly concerned with the benefits from health care.

**MAXIMISING PRINCIPLES**

According to maximising principles, justice requires that health care should be distributed so as to bring about the best possible consequences (although, of course, different maximising principles evaluate consequences in different ways). Maximising principles are popular among economists, since they are broadly utilitarian in focus (although rarely the same as classical utilitarianism) and fit well with the consequentialist logic of the classical model of rational choice that underpins standard economic theory. What such principles imply always depends crucially on what estimates are made about the likely consequences of different courses of action.

Perhaps the most obvious maximising principle in this context is to maximise aggregate population health. This is not the same thing as classical utilitarianism, since the consequence being valued is health rather than happiness (which presumably requires more than good health alone). In our rationing exercise, use of this principle would essentially imply treating the patient who is expected to gain the largest total amount of health over his or her remaining lifespan, although also taking into account possible indirect consequences for other people’s health (for example, due to health care resources freed up for treating other patients). The actual implication will depend crucially on what estimates are made about the expected health gain.

Our working assumption is that Steve gains the most health. However, if we modify this to assume that Daniel can be expected to live a full life if the treatment is successful, then Daniel would come out top. It is also possible that this implication might have to be modified if we estimate the indirect health consequences of the decision. One set of indirect consequences involves the costs of long term care for Marinder (both financial and in terms of a burden to her son), who will not be able to live independently without the hip replacement. The resources freed up by reducing the burden of long term care for Marinder might be used elsewhere to improve the health of others.

Another possible indirect consequence might be health gains to patients in the future from knowledge learned by trying out the experimental drug on Daniel (which would, of course, count in Daniel’s favour). A third is the possibility that Steve’s and Daniel’s parent(s) will suffer ill health if their son is denied treatment (counting in both Steve’s and Daniel’s favour).

A less obvious form of indirect health consequence is the possibility that, if successfully treated, Daniel might be able to have children. If Daniel’s unborn children are straightforwardly counted as an addition to the relevant population, this might swing the principle of maximising total population health back in his favour (since treatment is less likely to influence the ability of any of the other patients to have children). However, maximising principles are usually taken to refer to the existing population only, and adding people to the population is typically not counted as a benefit. Furthermore, the issue of valuing unborn generations raises a number of unresolved conceptual difficulties.

A rather broader maximising principle is to maximise wellbeing or flourishing, which includes aspects of wellbeing other than health. This is not necessarily the same thing as classical utilitarianism, since wellbeing can be understood in terms of “objective” capabilities (for example, the ability to form and achieve goals, to interact with others, and so on) as well as subjective pleasures or desires. In our example, considerations other than health gains might include wider (non-health) aspects of each patient’s quality of life, and how much anguish parents or relatives might feel. Given further information, either or both of these considerations might modify the health maximisation implication. In the absence of this information, however, we cannot tell whether or how the implication might change.

**NEED PRINCIPLES II: NEED AS CAPACITY TO BENEFIT**

So far, the need principles reviewed in this paper have all defined need in terms of ill health, more or less broadly understood. However, some economists have argued that any definition of need in terms of ill health is inadequate because it pays no attention to how much benefit the health care is likely to bring. No matter how ill a patient is, it is hard to see how she can “need” health care that does no good. Economists have proposed two alternative definitions of need which might be able to deal with this point.

The first proposal is to re-interpret need directly in terms of the individual’s capacity to gain health from treatment. In our rationing exercise, this interpretation would require us to rank patients in order of health gain, as a maximising principle would essentially imply (ignoring indirect health consequences). More generally, however, the principle “distribute in proportion to capacity to benefit” is not the same as the principle “maximise health”. A maximising principle will imply concentrating resources on those who stand to gain the most, possibly to the exclusion of those who gain the least. (A proportionate) need principle, by contrast, will always imply giving some health care to those with lesser needs (ie, in proportion to those needs).

In our example, however, a proportionality principle is hard to apply literally, because it will do no good to share the £4,000 between all four patients. The only way to apply it literally would be to hold a cleverly designed lottery in which the probability of being treated is proportional to capacity to benefit, thus maintaining proportionality in the distribution of expected health care.
Some economists have argued that identifying need with capacity to benefit implies a bias against people who need expensive treatments. If two people, A and B, have the same need (capacity to benefit), but treating A is more expensive than treating B, then distributing expenditure according to capacity to benefit might mean that B gets treated but A does not, even though they have the same needs. To incorporate this point, it has been proposed that need should be interpreted as the cost of treating the patient so as to exhaust capacity to benefit.  

In our exercise, if we assume that no further beneficial treatments are available to each patient, then a need principle based on this definition would imply giving equal priority to all four patients (since each treatment costs the same).

**Egalitarian Principles**

According to egalitarian principles, health care resources should be allocated so as to reduce inequalities in health. Most authors who advocate egalitarian principles would in fact not pursue equality as a sole objective, but would rather combine the goal of equality with other principles of justice (such as maximising health). We can still ask, however, what a "pure" egalitarian principle would recommend, as a first step before examining how further considerations might modify this recommendation.

One form of equalising principle is the "fair innings" argument that everyone is entitled to a similarly long and healthy life. A strict principle of equalising lifetime health, with no weight given to any other principle of justice, would perhaps give top priority to Daniel (who has the longest lifetime health expectancy), then Joanne, then Steve, and finally Mirander (who, being elderly, has already enjoyed the longest lifespan).

A second kind of equalising principle focuses on equalising people's **opportunity** for lifetime health, rather than achieved levels of health, in order to account for individual freedom of choice and autonomy in making choices that influence health. According to this view of equality, justice is done if people who choose not to exercise their opportunity for health (for example by leading an unhealthy lifestyle) achieve lower levels of health. If, for instance, Joanne's drug taking were to be regarded as a free choice, whereas Steve's car accident were regarded as beyond his own control, then Joanne's lower lifetime health expectancy might be considered just on an equality-of-opportunity view. This would then modify the recommendation of the "fair innings" principle, by giving less weight to Joanne and more to Steve. In the extreme, it might even be considered that Joanne had a greater opportunity for lifetime health than Mirander (for example, if she came from a more privileged background), and was entirely to blame in squandering this opportunity through drug abuse. The assumption that Joanne's predicament is entirely her own fault (and had nothing to do with socio-economic conditions or bad luck) seems implausible, however, and so the principle of equality of lifetime opportunity for health seems unlikely to imply giving Mirander higher priority than Joanne.

**Combination Principles**

Combination principles combine different principles together in a structured manner—ie, by clearly specifying how the combination is to be done. One way to combine principles is for a secondary principle to come into operation only when the primary principle does not yield a definite answer. Harris has argued for a combined principle of this kind, which combines a narrow (life-saving) needs principle with a lottery principle. He proposes that priority be accorded to saving life as the primary principle, and that a lottery should come into operation if two or more people have equal immediate threat to life and resources remain scarce.

A second way to combine principles is for two principles to be weighted together, neither having absolute priority over the other. Williams has advocated a quantifiable combination principle of this kind, which gives weight to both the equalising principle of reducing inequalities in lifetime health and to the maximising principle of maximising aggregate health. The ranking of the patients in the rationing exercise would depend on the relative weights given to these two principles (ie, society's degree of aversion to inequality). We have assumed mild aversion to inequality, so that the ranking is largely determined by health maximisation alone (with Steve and Joanne retaining first and last places, respectively) but Daniel gains priority over Marinder on grounds of reducing health inequalities.

Lockwood has also argued for a combination position, but this time one which combines a need principle with a maximising principle. More precisely, he endorses the need principle of distributing in proportion to immediate ill health, coupled with the maximising principle of maximising health. It is not clear what this would imply in our exercise, because it all depends on how much weight these conflicting principles are given, and Lockwood gives no indication of how this weighting should be done.

One way of putting into effect a combination principle of this kind so that it does yield specific implications has been explored by Hadorn, and was used in modified form by the Oregon Health Services Commission to draw up a package of health care services that should qualify for public funding. In essence, the procedure is to rank health care services according to their degree of relative effectiveness, and then to calculate how far the health care budget will stretch down that list of treatments. Only those health care services which are ranked above this line are funded, so that health care is provided only when its effectiveness is above a certain threshold. This procedure gives some weight to maximising health (because relatively ineffective health care is not funded) and some weight to distributing in proportion to capacity to benefit (because relatively effective health care is funded even if it is not cost-effective). If we applied
this principle to the budget of £4,000 in our example, then Steve would come out top of the list as his is the most effective treatment according to our working assumptions.

Broome has also proposed a combination of maximising and need principles.13 His proposal is that decisions should balance a maximising principle against the demands of “fairness”, which he argues should be understood in the context of health care as a need principle combined with a lottery in cases of equal need. However, Broome does not endorse a particular maximising principle or a particular interpretation of need; nor does he specify how the weighting of principles is to be done. So it is hard to know what his combination principle would recommend in our exercise.

Arguably, though, the demands of “fairness” do not come into play in our particular example, because it will do no good to any of the patients to share out the £4,000 among them in proportion to need. If so, Broome’s combination principle in this case would boil down to a straightforward maximising principle, which (as discussed above) would give priority to Steve.

Comparison with what the public thinks

A previous study has examined public views about the hypothetical rationing decision we have been using to illustrate alternative rationing principles.1 This study involved 60 members of the general public drawn from two urban general practices in the York area of England. Respondents took part in two separate small-group discussions of various questions about health care rationing, including the hypothetical rationing decision discussed in this paper. Respondents were told that the purpose of the hypothetical exercise was to encourage discussion of general ethical principles, and that it was not intended to be a realistic example. The detailed methodology and results from this study are presented elsewhere.14

The basic finding of this study was that the public support a combination of all three kinds of substantive principle proposed in the academic literature: needs, maximising and egalitarian. Daniel was the top or joint-top priority for 80% of respondents, followed by Marinder and then Steve, with Joanne receiving lowest priority on average (although with considerable variation between individual respondents in their second and lower-level rankings). The most common justifications for the decisions were that: i) Daniel, Marinder and Steve gain more health from treatment than Joanne; ii) Daniel is a child; iii) Daniel is in urgent need of life-saving treatment; and iv) Joanne’s illness is self-inflicted. There was near universal agreement that the first three of these considerations should be taken into account, but considerable debate and disagreement about the fourth. It was clear that no single consideration had overriding weight, but rather that respondents weighed these sometimes conflicting considerations together.

It is remarkable that, although all of these principles have been discussed in the literature in isolation, to our knowledge this combination of principles which seems to be favoured by the public has never been proposed in the literature or developed into a coherent theoretical position.

Discussion

Three main types of principle for justice in making health care rationing decisions are commonly proposed in the academic literature: need principles, maximising principles and egalitarian principles. However, each of these principles is more narrowly focused than the pluralistic combination of principles which appear to be supported by the general public, and philosophers have yet to develop a coherent theory which combines all three principles.

Of course, one possible reason why philosophers have not developed a theory of justice that accords with public opinion is that they do not see this as their job. Instead, they may see it as their job to point out the errors and confusions of popular thinking. In other words, the public may simply be wrong or confused in trying to combine such a pluralistic combination of rationing principles.

However, a second possible reason is that philosophers have generally been reluctant to address the issue of justice in health care rationing at all. Much of the literature on this topic (but by no means all of it) is written by health economists. Philosophers specialising in medical ethics have tended to concentrate on decisions about particular patients rather than decisions about the distribution of health care among patients. And those specialising in moral and political philosophy more generally have tended to concentrate on constitutional decisions about the basic institutions of the health care system, rather than policy decisions about the day-to-day running of that health system.

This reluctance of professional philosophers to tackle day-to-day policy questions about justice in health care rationing can be seen in the work of Daniels,22 who has extended Rawls’s famous theory of justice23 to apply to the health care system. Daniels advocates a general theory according to which the health care system should be designed so as to bring everyone as close as possible to a decent minimum level of health, which he refers to as “normal species functioning”.24 But, he acknowledges that policy-makers will not find answers in his theory to “their most immediate and pressing questions”.25 This same reluctance is also demonstrated by Richard Hare who advocates a sophisticated utilitarian theory and has applied it to a variety of problems in applied ethics.26 In the context of health care, however, he says that: “I am not very well versed in the details of these problems, and have had to content myself with giving what I think is the best philosophical basis for their solution, leaving it to others to apply it to the various difficulties that all who seek to provide an adequate health service are faced with”.27

In conclusion, then, it is our contention that there is scope for further philosophical work in developing new theories of justice in health care.
rationing, which are more pluralistic than distribution according to need, or health maximisation, and more consonant with the commonsense moral intuitions of the public.

Acknowledgements
We would like to thank Nottingham Health Authority for permission to use their “How would you choose?” exercise. We would also like to thank John Broome, Tony Culyer, John Harris, Paul Menzel, Peter Singer and Alan Williams for extremely helpful comments, although responsibility for the content of this paper lies entirely with the authors. This research was partly funded by the Northern and Yorkshire Regional Research and Development Office, whom we would also like to thank.

Richard Cookson is Senior Lecturer, the Health Economics Group, School of Health Policy and Practice, University of East Anglia. Paul Dolan is Reader in Health Economics, School of Health and Related Research and Department of Economics, University of Sheffield.

References
7 See reference 3: 71-5.
18 Culyer AJ. Need—is a consensus possible? Journal of Medical Ethics 1998;24:77-80.
21 See reference 16: Lockwood M. Quality of life and resource allocation.