
Chapter 17

Allocation of Scarce Resources

Paul Menzel

Resources, including health care resources, are almost always scarce, and they somehow have to be allocated. Often, though, we do not perceive them as either scarce or allocated, particularly when no person is making conscious decisions about how those resources get distributed. Instead some structure in which we work or live just handles the distribution without us having to think about it. When in the US, for example, a much greater volume of health care services is delivered to well-insured citizens than to those who are un- or under-insured, it is the structure of health insurance and financing that explains much about how the allocation comes about, without anyone making “decisions” to allocate more to the well-insured than the relatively uninsured.

A distribution like this could perhaps be regarded as still an “allocation” of health care, but the usual sense of that term involves more conscious decisions by particular persons or groups of people who direct different amounts of care to different people. This entry will focus only on allocation in this latter sense: the relatively deliberate, conscious distribution of care to different persons. Even among such consciously distributed care, however, we should notice a large segment that we still would not count as “allocated.” Competent health care professionals are always consciously aiming care differentially at different patients, but that is just “good medicine.” A physician’s primary responsibility is to get medically appropriate care to the patients who need it for their condition and not to those for whom it is inappropriate. Accordingly, I will use “allocation” in this entry only to encompass care consciously distributed among a population of citizens and patients for reasons that go beyond professionally determined medical need.

As much as is thus eliminated from the scope of “allocation,” however, much remains. In any situation of scarcity, resources are insufficient to handle all medical need. Flu vaccine ends up in short supply, for example, and someone has to determine an order of priority among all the potential patients who want or need it. Should infants, who are most susceptible to influenza but will seldom die from it, get the vaccine first, or should it go to the very elderly who, though they contract influenza at a lower rate, are more likely to die from it? Allocation of this sort is

often called “rationing,” but that term’s harsh sound to many ears can perhaps be skirted by using the less provocative but still transparent term, “prioritizing.”

Micro- v. Macro-allocation, and a Quandary for Clinical Practice

Within such a meaning of allocation, bioethicists, health policy analysts, and health economists typically distinguish between micro- and macro-allocation to signify different contexts in which prioritizing decisions are made. Micro-allocations occur within a specific delivery setting rather than more population-wide; physicians in a given hospital, for example, have only so many renal dialysis beds, and more patients are referred to the unit than it can accommodate. Macro-allocations of dialysis, on the other hand, occur at the level of a regional or state coordinating agency, or in a report making recommendations about prioritizing care for potential dialysis recipients. (Note, of course, that structural elements in a situation like this may hide the fact that anything is allocated. If some kind of insurance ends up covering virtually all prescribed renal dialysis – as, e.g., Medicare in the US does – then supply may expand to accommodate demand without any further conscious decision being made to “allocate” anything. Resources are still being allocated to dialysis as a category of care, of course, in society’s “act” of having Medicare cover dialysis.)

One important controversy about how and by whom prioritizing and allocating should be done can be seen as an argument about the relationship between micro- and macro-allocation: should a physician or nurse ever engage in “bedside rationing”? Such decisions by clinicians would seem to be a violation of their oath to serve the welfare and dignity of the patient they are treating, for in “bedside rationing” providers consider the relative benefits of care that might be given to whole different categories of patients before deciding whether the particular patient before them warrants provision. The issue becomes whether an intrusion of macro-level considerations into the decisions of clinicians constitutes an abdication of providers’ moral duties to their patients. This question has become pervasive in an age where managed care is managed partly precisely to control costs. On the one hand, patients come to providers trusting that providers’ professional fidelity is to them. On the other hand, aren’t patients members of the insurance pools that have to wrestle with difficult prioritizing questions about the pools’ scarce resources? Can patients legitimately expect providers not to be, at least somewhat, bedside rationers (Pellegrino and Thomasma, 1988; Menzel, 1990; Morreim, 1991; Ubel, 2000)?

Some examples of the strong moral pull in both directions of this dilemma are striking. Take the use of a temporary artificial heart or ventricular assist device (VAD) as bridge to a hoped-for transplant for a patient on an organ waiting list. Understandably, given that the patient is at death’s doorstep and has only one route to significantly longer life – a transplant – the bridge device is *urgent and necessary rescue care* for that individual. On the other hand, it is absolutely clear

that given the real shortage of human organs for transplant, the *net* lifesaving capacity of such bridge devices is *zero*. They only shuffle people around on the list. Instead of dying soon and allowing someone else to move up on the waiting list, the person using the device tries to find an organ, and if she does so, it will almost certainly be one that would otherwise have gone to someone else. But if bridge devices thus save no lives (net) whatsoever in the larger picture of the organ-scarce transplant world, how can one defend their use and expense? Doctors who continue to prescribe such devices are providing the world *no net lifesaving at all*.

Before leaving the bedside rationing issue with the impression that it constitutes simply a knock-down, drag-out tussle between a provider's loyalty to the individual patient and consideration of the larger total good of all patients, we should recognize that "reconciliationist" views may be plausible, claiming that a wiser and more efficient use of resources can be reconciled with the moral obligations of practitioners to their individual patients. Reconciliationists come in at least three stripes.

- 1 *Separation of roles.* Parties more distant from the patient than the immediate clinician should make prioritizing decisions by constructing "practice guidelines." Clinicians should then ration only within those pre-determined guidelines.
- 2 *Patients as larger, autonomous persons.* The persons who are patients are not just patients. They are also subscribers and earlier patients, and bedside rationing can be morally grounded in the consent of the patient at an earlier time to restrictions on his or her later care. In the bridge-to-transplant case, for example, imagine patients who have just come onto the transplant waiting list. If asked whether they are willing to adopt a policy of abstaining from the use of any bridge devices, won't all of them likely say that they are, once they realize that (a) their own chances of being saved by a transplant are just as high if no such devices are allowed as their chances would be were such devices allowed, and (b) resources would be "left over" for some other value-producing use (Menzel, 1990; Hall, 1997)?
- 3 *The patient as a member of a just society.* A provider's proper loyalty to a patient, though it is not directly dictated by efficiency, is to the patient as a member of a just society. That loyalty enables the clinician to prioritize and ration with a clean conscience by basing her decisions on considerations of fairness and justice (Brennan, 1991).

Such reconciliationist views may seem an ideal way to handle the conflict between clinician loyalty to patients and societally efficient resource use, still allowing conscientious clinicians to play a major part in allocating scarce resources. These views may face great difficulties in actual use, however. The separate-roles view will involve formulating detailed, care-rationing practice guidelines in abstraction from some of the medically relevant particulars of individual patients. By contrast, more active bedside rationing in which clinicians make substantive rationing decisions may be more efficient and more artful, and therefore preferable (Ubel, 2000).

The prior-consent-of-patients view requires not only accurate readings of what restrictions people are actually willing to bind themselves to beforehand but also a willingness of subscribers and citizens to think seriously about resource trade-offs and then abide honestly by the conclusions of their prior thinking even when that places them on the short end of rationing's stick. And the patient-in-a-just-society model requires more agreement on what constitutes a just society than is possible.

Allocation and Theories of Justice

Health care systems are sometimes thought to fall on a spectrum from those that are characterized by great equity of distribution but employ considerable collective coercion to those that achieve little equity but encroach much less on individual liberty (Culyer et al., 1981). Usually systems lying in the equity-emphasizing and arguably liberty-diminishing direction are also surmised to be likely more efficient (cost-effective). More centralized, they seem better able to control costly provider behavior and insurance-distorted patient demand. Their structural paradigm is a unitary public system, either single payer insurance or national health service, and justice is usually taken to be the dominant value that they serve. The paradigm on the other, decentralized end of the spectrum is a pluralistic system of market competition that does not achieve universal access to even a basic minimum of services and is therefore regarded as leaving allocation to considerations quite other than those of justice.

Two cautions are in order, however, about these preconceived associations of various structural options for the delivery of health care with differing emphases on justice, liberty, and cost-effectiveness. First, it is highly debatable whether the stereotypical conflicts of liberty and cost-effectiveness with equity should be regarded as conflicts of liberty and cost-effectiveness with *justice*. Senses of justice vary widely across the political spectrum, and libertarians and utilitarians as well as contractarians and egalitarians have their own, different views of justice. Utilitarian theories of just allocation of scarce medical resources are likely to focus significantly on services' cost-effectiveness and costworthiness. Contractarian and egalitarian views of justice generally insist from the start on both universal access to a robust basic minimum of care and financially equitable burdens of financing (for example, community rated premiums), and they emphasize that allocations of resources in situations of scarcity must respect the equal value and dignity of every individual. Libertarians do not just speak of denying claims of justice in favor of claims of liberty; they think that justice itself requires a high respect for liberty and the pluralistic market that it generates (Engelhardt, 1996).

Second, conflict between the values of individual liberty and responsibility with the values of both equity and cost-efficiency may be less intractable than is usually assumed. If it is, the result will be to break the rigid and stereotypical association of senses of justice with the two paradigms of basic organizational structure for a health care system: libertarian-leaning senses of justice with pluralistic market

systems that employ little central coercion, and egalitarian senses with unitary public systems. Critically examining and understanding better the ingredients of liberty, equity, and justice as well as the complexity of how they actually intersect in a health care system may open up different possible associations.

Here, for example, is one such different view (Menzel, 2002). Even non-egalitarian, libertarian leaning, responsibility-emphasizing views of distributive justice, this view claims, should strongly embrace compulsory, universal coverage of health care for basic care. A primary reason is the high degree of objectionable, irresponsible free-riding that occurs in a voluntary insurance system; those who do not insure (or pay to get their employees insured, for example) usually get basic care anyhow, paid for by cost-shifting that raises prices and premiums for those who are insured. In a complementary concession from the opposite end of the political spectrum, those who hold egalitarian views of justice ought not to regard different levels of health care coverage for people of different income levels as necessarily unjust. Such variety may reflect the varying relative values that differently situated persons put on higher and lower consumption of health care, and thus a lean system of universal access to only very basic care may be just and equitable.

Cost–Utility Analysis as a Framework for Allocation

Health economists have developed a number of models for allocating care that aim at achieving efficiency – the most health benefit for the least cost. Efficiency is not just some mysterious gospel of economists. It seems commonsensical: why not see the point of prioritizing health services to be the same as the point of medicine itself – namely, to create greater health? The intuitive attraction of economic efficiency is rooted in its basic concept of “opportunity cost”: the value sacrificed by not pursuing alternatives that might have been pursued with the same resources. When the value of any alternative use is less than the value of the current use, the current one is efficient; when the value of some alternative is greater, the current use is inefficient.

In thinking of the possible alternative uses, our sights can be set either narrowly or broadly. If we focus just on other options in *health care*, wondering whether we can get *more benefit* for our given health-care dollars, or whether we can get the *same health benefit more cheaply*, we are engaged in cost-effectiveness analysis (CEA). If, on the other hand, we are comparing an investment in health care with *all the other things* we might have done with the same time, effort, and money, we are engaged in cost–benefit analysis (CBA). CEA asks whether the money spent on a particular program or course of treatment could produce healthier or longer lives if it were spent on other forms of care. CBA involves an even more difficult query: whether the money we spend on a particular portion of health care is “matched” by the benefit. We determine that by asking in turn whether, spent elsewhere, it could produce greater value of another sort, not just healthier or longer lives.

Both kinds of analysis are important in allocation decision-making, though CBA has been less developed by health economists than CEA. CEA – getting the most health and life for a given size investment in health care – seems more feasible and less ambitious: compare different health-care services, detecting either final differences in expense to achieve the same health benefit or differences in some health benefit (added years of life, reductions in morbidity, etc.). We don't have to compare health benefits with opportunity costs having nothing to do with health. Yet even CEA – achieving efficiency *within* health care – may be very difficult. How do we really compare the values of non-life-extending hip replacement, for example, and life-extending dialysis or transplantation?

A major step taken by health economists at this point is to develop the notion of a unit of “health utility” (or “health state utility”), a common valuational currency for all losses and gains in health. The specific conceptual units developed to represent this go by various labels. Two examples are a “well-year” and a “quality-adjusted life year” (QALY, pronounced to rhyme with “holly”). The idea is to construct or discern a unit that combines mortality with quality of life considerations – “a year of healthy life” (Williams, 1985). We can then compare not only life-prolonging measures with each other but also measures that enhance quality with those that prolong life – hip replacements with kidney dialysis, for example. And then we can also track the health of a population, calculating changes in per capita “years of healthy life.” In both cases, we will be doing “cost–utility analysis” (CUA).

Having available a unit of health utility that combines mortality and morbidity will be immensely useful if we are trying to maximize the “health benefit” of a given amount of resources invested in health care. Suppose dialysis patients' self-stated quality of life is 0.8 (where 0 is death and 1.0 is normal healthy life). They would gain 8.0 QALYs from 10 years on \$40,000-a-year dialysis, a cost–benefit ratio of \$50,000 per QALY. Suppose hip replacements improve 15 years of life from 0.9 quality ranking to 0.99. That will be a 1.35 QALY gain for the \$10,000 operation, a cost of less than \$7,500 per QALY. To achieve greater efficiency, we apparently should expand the use of hip replacements and look toward reducing dialysis. Similar CUA's of other health services can be done using the basic elements of size of quality improvement produced by treatment, duration of that improvement, and the number of persons who gain it.

A sizable literature of CUA has developed, not only studies of particular procedures but also intense discussions about how to construct a common unit of health benefit. Take the QALY. Questions abound. Whom does one ask to discern quality-of-life rankings for different sorts of health states – patients with the problems, or other citizens and subscribers who are less dominated by their desire to escape their immediate health need? What questions do we ask them? Those building the QALY and well-year frameworks have used “time trade-off” (how much shorter a life in good health would you still find preferable to a longer lifetime with the disability or distress you are ranking?), “standard gamble” (what risk of death would you accept in return for being assured that if you did survive, you would be

entirely cured?), and several others. Whatever question people are asked, it should convey as accurately as possible what might be called the “QALY bargain”: their exposure to a greater risk of being allowed to die should they have an incurable, low-ranking condition, in return for a better chance of being helped to significant recovery or saved for prospectively normal health.

The moral argument for employing CUA and using some common health benefit unit like the QALY is more than just a narrow focus on aggregate economic efficiency per se. A major argument for using both quality adjustment as well as longevity extension in a serious attempt to maximize the benefit that a plan or an entire health care system produces is that it is people themselves who implicitly quality-rank their own lives and thus consent to the allocation priorities that QALYs or well-years generate.

Critics charge, however, that maximizing years of healthy life in our lifesaving policies systematically fails to respect the individual with an admittedly lower quality of life (Harris, 1985). To what interpersonal trade-offs have people consented, when those might involve themselves? Suppose you yourself prefer a shorter, healthier life to a longer, less healthy one. You have now become, let us say, an accident victim who can, if saved, survive only as a paraplegic, while someone else can be saved for more complete recovery. Admittedly, you yourself prefer a life with recovery to one with paraplegia, and you would be willing to take a significant risk of dying from a therapy that promised significant recovery if it succeeded. But here is the problem: you do not admit – and you never have admitted – that, when life itself is on the line, a life with paraplegia is any less valuable to the person whose life it is than life without that disability is to someone else. Compared with death, *your paraplegic life would still likely be as valuable to you as anyone else’s “better” life is to them.* That is, you want to go on living as fervently as the non-disabled person does. Conventional CUA thus raises significant questions about discrimination against the disabled and chronically ill when it is used to prioritize among various opportunities to save lives. (This consideration will surface again in the Rule of Rescue segment of the next section.)

Accounting for Specific Factors Other than Health-Related Utility

If we have thus opened up the search for *reasons* to use CUA in making allocation decisions, we may already have departed from a purely utilitarian framework within which to make allocation decisions. Not surprisingly, conventional CEA – that is, CUA – provokes many of the same ethical objections that plague utilitarianism in general: inadequate attention to the individual person in relationship to the aggregate (“common”) good, and insensitivity to issues of distributive justice involving those who are less advantaged. Some will even just dismiss CUA because of the alleged weaknesses of the utilitarian philosophy that it reflects. Others have urged, however, an alternative, usually termed “cost–value analysis” (CVA):

examine specific values that CUA fails to incorporate, with an eye toward reforming the methodology of CEA itself into something other than CUA (Nord, 1999 and Nord et al., 1999).

Included among the arguably non-utilitarian factors that CUA currently underestimates are patient age (age itself, as distinct from the association of age with duration of effect), the initial severity of illness from which a treatment produces an improvement, and patients' limited potential for increased health as a long-term identifying characteristic of their lives. Consequently, important social values of justice and non-discrimination can stand at odds with recommendations generated by CUA. Regardless of whether such arguably non-utilitarian social values could be incorporated into a revised model of CEA, they are important factors politically to consider in allocation decisions. What politician wants to defend policies which expose individual people to injustice in the name of an impersonal, aggregate good?

Although some of these factors may affect the "individual utility" dimension of value, they are mainly relevant when "societal value" is being calculated. The latter is a different kind of value. In societal value the focus is explicitly on *interpersonal* trade-offs – decisions about what services to provide among the wide array of possible services that often affect *different* groups of people. Typically such trade-offs are at issue in allocating preventive or acute care services at the "budget" or "coverage" level for large populations, as distinct from the "admission" or "bedside" level where such interpersonal trade-offs may be more debatable.

Whether as factors transcending cost–utility analysis that health economics might incorporate into their frameworks for allocation or as independent elements having little if anything to do with health economics, several variable characteristics of all the different people who have medical need are arguably plausible grounds for allocating scarce resources. Six will be discussed here: age, severity of illness, urgency of rescue, health potential upon recovery, maintenance of hope and assurance of treatment, and comparative duration of benefit (regarding these and others, see Menzel et al., 1999, and Kilner, 1990).

Age

While through duration of effect age is indirectly accounted for in CUA, age per se is not. Indirectly it is, as treatments for the relatively elderly often produce benefits of shorter duration than treatments for the relatively young, thereby yielding priority for the young. In addition to this indirect effect of age, however, there is notable evidence that for comparisons in which duration of gain is held equal while age is varied, most people express preferences considerably at odds with CUA's disregard of age per se. For example, many people would not only agree that more than one person would have to be saved for an additional 20 years to equal the value of saving five people (of the same age as the one) for an additional four years

each. Many would also say that more than one 70-year-old, for example, would have to be saved for 10 years of additional life to equal the value of saving one 30-year-old for an extra 10 years. Moreover, this preference continues when one rids respondents of any implicit presumptions that quality of life between 70 and 80 must be lower than that between 30 and 40 (Johannesson and Johannesson, 1996; Nord et al., 1996).

The moral ground for this preference for the young is undoubtedly thus egalitarian, not – or not only – utilitarian. The young should be treated first, even when their gain is smaller, so that they obtain a fairer and more equal chance of living a long life. This has been called the “fair go” or “fair innings” argument (Harris, 1985; Williams, 1997). There may, of course, be societies in which the elderly are so revered that these egalitarian considerations for at least some favoring of the young get overridden. In those contexts, CUA-recommended allocations that effectively ignore age per se may not be out of step with public preferences. For societies that do not have such strong differential reverence for the elderly, however, those allocations will be.

Societal values that favor the young, whether they be based on egalitarian fairness reasons or on utilitarian calculation, may give rise to charges of “agism.” Perhaps, in fact, the societal issues are so contentious that empirical investigation will show little in the way of any predominant social value toward greater priority for the young than is already indirectly accorded by duration-of-benefit considerations. In any case, however, the conventional utilitarian treatment of age – assuming that age is relevant only through its highly variable effect on the duration of treatment benefit – is questionable on both moral and empirical grounds.

Severity of illness

Utilitarian analysis incorporates the severity of the illness that a procedure treats only insofar as it is one of the two variables needed for determining the size of the treatment effect (initial health and treatment end point). It does not accord any weight to the severity of illness per se – the initial health-related quality of life. Both public official statements and population preference data, however, suggest that people often wish to give greater priority to those who are worse off before treatment, above and beyond any priority they may achieve by health benefit calculation (Nord, 1993a; Core Services Committee of New Zealand, 1994; Ubel, DeKay, et al., 1996; Ubel, Scanlon, et al., 1996; Nord, 1999).

For example, in one study of 150 Norwegian politicians accountable for health policy at the county level, subjects were presented with a dilemma: provide treatments for a severe illness A that would bring “a little” help, or treatments for a moderate illness B that would help “considerably.” They had three choices: divide resources evenly between the two illnesses and their treatments, allocate most to illness A (the most severe), or allocate most to B. Nearly half (45 percent) chose

equal division, and 37 percent chose priority for A. Only 11 percent chose priority for B (Nord, 1993a).

The key to severity's ethical relevance is captured by the fact that treating the more severely ill is helping those who are in greater need. Notable treatises in social philosophy have articulated a fundamental general priority for helping "the worst off" – those whose life prospects are the most disadvantaged (Rawls, 1971; Daniels, 1985). We reduce inequality if we give priority to patients with more severe illness. For example, if two treatments can raise one person's health-related quality of life from 0.5 to 0.9 and another's from 0.3 to 0.6, treating the less severe illness leaves a greater difference between the two persons' health states (0.3 and 0.9). Treating the more severe illness reduces inequality of health and leaves the two individuals with more nearly equal health quality rankings of 0.5 and 0.6.

Preference data suggesting an independent concern for severity would thus appear to have an ethical basis. To be sure, critics might challenge the data itself. Did the study respondents really understand the hypothesized treatment effect to be smaller for those whose initial illness is more severe? Perhaps respondents accord extra priority to the most severely ill not because they see them as more ill, but because they believe that the increments in individual utility obtained from treatment are actually larger than those accruing to those whose baseline health is better. Some studies, however, have used examples that conveyed particularly clearly that the more severely ill were benefiting from smaller treatment effects; it, too, yielded distinct preference for giving priority to those with the more severe illnesses (Nord, 1993a).

Lifesaving and treatment in the face of death – the "Rule of Rescue"

The most severe illnesses, of course, put people face to face with death. The propensity to regard situations where identifiable patients face great risk of avoidable death as holding a unique call on resources has been called the "Rule of Rescue" (Jonsen, 1986). Rooted in the Kantian tradition of considering the individual to whom one is relating as an ultimate end-in-herself, this "rule" resists the usual maximize-benefit calculation. Critics of conventional CEA such as Hadorn, for example, have argued that "any plan to distribute health care services must take [this Rule of Rescue aspect of] human nature into account if the plan is to be acceptable to society" (Hadorn, 1991).

The actual actions that people take provide ample evidence that we will expend great effort and large resources to avert death (the girl down the well, astronauts in space, sailors lost at sea, etc.). We also have examples from public life such as the state of Oregon, where all lifesaving services rose to a separate high priority category in the state Medicaid plan's eventual rationing list after having been treated in the conventional efficiency manner earlier (Hadorn, 1991).

It is possible, of course, that such disproportionate investments in immediate

lifesaving are irrational, the result of a weakness in our mental judgment that suppresses the greater aggregate gain in quality improvement or lifesaving that results if the same total investment is made in a larger number of less urgent but still health threatening situations. Compare, for example, saving two lives by eliminating 1: 50 risks of death for 100 people with saving one life by trying to save two people threatened by certain death if we do nothing but who each have a 50 percent chance of recovery if we act. The former action will likely save two lives, the latter action only one. The latter, though, takes place in a “death’s doorstep” situation. It is in fact irrational to invest in the latter before the former *if* the value of a life that is discerned by willingness to reduce only a statistical risk can properly be equated with the value of a life saved on death’s doorstep. But can it be?

This is, in fact, the problem at the heart of all claims to place a monetary or resource value on life based on people’s distinctly limited willingness to sacrifice to reduce moderate or low risks. Of what relevance are differences in perspective by degree of risk? One critic of claims to discern a limited monetary value of life has argued that in principle only valuations of life made directly in the face of death are correct reflections of the actual economic value of life (Broome, 1982). Another contributor to this discussion has noted that he does not know of anyone “who would honestly agree to accept any sum of money to enter a gamble in which, if at the first toss of a coin it came down heads, he would be summarily executed” (Mishan, 1985, pp. 159–60). Some conclude from this that there is no rational limit on what to do or spend to save a life because no particular finite amount of effort or money is adequate to represent the real value of life.

Even if this point about the actual value of a life is correct, however, it may not render estimates of a limited monetary value of life irrelevant for use in health policy. In the context of setting policy about whether to include a certain service in our package of insurance, we cannot just assume that the later perspective of an individual immediately in the face of death is the correct one from which to make decisions. Such a perspective may be proper for the legal system to adopt in awarding compensation for wrongful death, for there we are trying to compensate people for losses actually incurred. Arguably, however, health-care decisions ought to be made from an earlier perspective. In modern medical economies most people either subscribe to private insurance plans or are covered by public ones. Once insured, whether in private or public arrangements, subscribers and patients as well as providers find themselves with strong incentive to overuse care and underestimate opportunity costs. Why should we not address the problem of controlling the use of care in the face of these value-distorting incentives at the point in the decision process, *insuring*, where the major cost-expansion pressure starts? In health policy, while it may not be necessary to claim that willingness to risk life shows us the “value of life,” willingness to risk may still be appropriate to use.

In any case, regardless of these basic considerations for and against thinking in terms of a monetary value of life, we need to keep in view the possibly broader context of any special value of lifesaving. The special value of treatment in the face of death involved here may pertain to more than lifesaving services. Widespread

attitudes toward hospice and other non-lifesaving terminal care also suggest the same special value of care in the face of death. Putting up with severe pain for a six-month period when one expects to live for many years is one thing; having to put up with it at the end of one's life is another. People generally, not just patients facing death, have a special concern that life not end in pain. Thus, palliative measures for patients with terminal conditions may produce an extra value. Perhaps the "Rule of Rescue" is more appropriately expressed as the special value of "Care in the Face of Death."

Level of health potential

The societal value of priority for treatment of those with more severe illnesses focuses independently on a patient's start point, as distinct from the size of treatment effect. Analogously, the end point may have independent relevance that is not accounted for in a utilitarian calculation of the size of treatment effect. Call this factor the "level of potential." The social value it expresses is a reluctance to discriminate against patients who happen to have lower potential for overall health, or do anything that would reduce the priority given to their care.

Suppose that on a subjective utility scale, treatment can move one person from 0.6 to 0.8 and another from 0.6 to 1.0, and that the first person's end point of 0.8 represents her maximum prospective health potential. Should we really regard the second person's treatment effect as having twice the value of the first person's? Treatment can, after all, "fully cure" the first within the perspective of her life. Her health potential defines the life she can lead. Since a life with that health potential is the only life that she will in any case have, it would seem plausible to say that reaching her 0.8 level counts as more than half the value of another person's move from 0.6 to full health. The essential ethical claim here is that where people are "located" in life in relation to their realistic potential is an important factor to take into account.

The same study of Norwegian politicians involved in health care decision-making previously mentioned in connection with severity of illness, also provides suggestive empirical support for this level-of-potential factor (Nord, 1993a). Perhaps the empirical data confirming consideration of returning to a previous level of limited health potential as an important societal value are less clear than those available to confirm severity of illness or urgent lifesaving as prioritizing factors, but they still point to societal values that conflict with conventional CUA (Nord, 1993b).

A relatively simple thought experiment confirms this intuitive power of an individual's maximum health potential against the results of conventional utilitarian calculation. Imagine two groups of patients stricken with a life-threatening illness. The first group were previously in full health and can be returned to full health with treatment. The second group previously had paraplegia and, with treatment

of their life-threatening condition, will continue to have it. Both, if treated, will live the same number of additional years. Assume that the health-related quality of life with paraplegia is 0.8, as calculated from “time trade-off” responses in which persons with paraplegia themselves expressed a willingness to sacrifice 20 percent of their remaining life extension to obtain a complete cure of their condition. CUA would recommend priority for saving the first group to full health and would shift to priority for the paraplegia group only if the number of lives saved there for the same total cost was at least 25 percent greater than in the other group saved to normal health. Yet few among us, reflecting seriously about the value of continuing to live, honestly believe that it is less important for society to save the lives of persons with paraplegia than the lives of others to full health (Nord, 1993b). The equal value of lifesaving regardless of final health state appears to dominate.

These societal preferences to save lives regardless of total health benefit may be based in more individual reasoning, too, and arguably they reveal special concerns about discrimination against the disabled and the chronically ill (Hadorn, 1992). Suppose, again, that the disabled person has ranked her individual quality of life at 0.8. This willingness to accept a 20 percent shorter remaining life in order to be cured from a permanent disability does not in any way indicate that she thought that her *life*, in relation to the prospect of *death*, was any less valuable and important to save than *another* fully healthy person’s life (Menzel, 1990, p. 84). The 0.8 that expresses willingness to trade time *within* a life does not constitute a judgment about the relative value of saving *different persons’* lives.

Note that the independent relevance for both the end and the start points of treatment – level of potential and severity of illness – above and beyond their role in determining the size of treatment effect, does not require us to dismiss the importance of size of treatment effect. It is only that gains in health-related quality of life need supplementing by independent consideration of severity of illness and level of potential.

Maintenance of hope and assurance of treatment

Conventional utilitarian analysis views societal benefit as directly proportional to the combination of the average improvement in quality of life and the number of people benefited. Two other factors, however, may qualify this simple linear relationship between numbers of people and total value.

One of them is “maintenance of hope.” Suppose we can allocate resources to one program, administered to a smaller segment within a larger group, that will yield a certain level of beneficial outcome for a finally greater number of beneficiaries, or to another program that generates that same individual outcome for what turns out to be a smaller number of actual beneficiaries though the segment of people *treated* is larger. The efficient allocation would be to devote *all* of the available pool of resources to the former program. Several studies, however, reveal that

many people would wish to preserve for each individual the hope that they still may receive treatment; to accomplish this, they prefer use some resources inefficiently (Ubel and Loewenstein, 1995).

Another factor disrupting the linear relationship between numbers of people and total value is “assurance of treatment.” Many people prefer extending entitlement for a treatment from some to all in a disease or treatment category even when such extension is inefficient (Ubel, DeKay, et al., 1996; Ubel, Scanlon, et al, 1996). They consider it inequitable to exclude some from treatment that most others receive, regardless of the increase in cost per benefit when treatment is extended to all. To satisfy desires for assurance of treatment, we might offer a treatment to everyone in an illness category despite the diminished odds of cost-effective outcomes for many of those added at the margin. In either case, we will have contradicted the outcome-efficiency orientation of conventional economic analysis.

Duration discounting

Conventional calculation of value achieved takes duration of health effect into account by multiplying the value of one year of a treatment effect by the number of years for which the effect lasts, usually with one important additional step: each year past the first is discounted for time preference back to its present value. Empirical studies, however, suggest a different pattern in people’s valuations of the importance of duration – perhaps a different pattern when they are evaluating different potential durations in their own lives, and very likely a different pattern when they consider duration in the context of interpersonal comparisons.

In the latter context, for example, respondents in Australia thought that saving 10 people for 10 years each would be equivalent to saving 7 people for 20 years each (Nord et al., 1996). That is, it took 140 total years accumulated in 20-year spans to equal in value 100 total years accumulated in 10-year spans, a reduction of 57 percent in the average value of a year of life in the last 10 years of each 20-year span (Menzel et al., 1999).

There are at least three possible reasons for such discounting beyond the pure time-preference used in conventional economic analysis.

- a At the level of individual utility, there may be a “quantity-effect” similar to the phenomenon of diminishing marginal utility generally: because of the benefits already achieved, the last year of life in a ten-year span of experience is not seen to be as valuable as any one of the earlier years.
- b At the level of societal value, people may have an aversion to inequality that leads them to value duration less than proportionately. It seems more equitable, for example, to extend two persons’ lives for 10 years each than one person’s life for 20.

- c Also at the level of societal value, an element may operate that is similar to the ethical basis of the level-of-potential factor: patients' shorter life potential should not be held against them.

Regardless of what explanation dominates, it appears safe to say that to adequately capture people's values, allocation decision-makers may need to discount the value of longer spans of life or health gained more, and perhaps considerably more, than is already occurring as a result of what economists call people's "pure time preference." Health policy and health economics need to embark on a wide range of future studies of duration, keeping pure time preference of the individual distinct from the other elements which may arise in social values. Such empirical research will hopefully reveal the relative importance of these different elements in people's discounting of the value of longer durations of health benefit. Moreover, a possible outcome from such research is the acceptance of the relevance of duration discounting for some reasons but its rejection for others.

Other influential factors

Undoubtedly there are other factors that can arguably be considered relevant to allocating scarce health care resources. Among them are personal responsibility, citizenship and nationality, compensation for those who are generally socially disadvantaged, health effects on others, non-health effects on others, the degree of either personal or community control seen to be preserved by a given program. They are only noted here, without further articulation.

Further Questions and a Concluding Note

Numerous other methodological questions arise in the struggle to make justified allocations of scarce health care resources. A few of them are:

- What group do we ask to obtain the most applicable values or preferences on any of the various factors discussed in the previous section? Patients with the conditions being evaluated, representatives of the general public or insurance plan members asked to imagine themselves to have the conditions related to the trade-offs being examined, or some other group such as politicians or health care providers?
- Should the aggregate, amalgamated preferences of a society derived from some democratic process dominate allocation decisions? Should those be allowed to effectively swamp out the preferences and values of the individuals who are most affected by prioritizing decisions?

- Should the values that ought to guide prioritizing decisions be thought of as empirical preferences that people in a given society happen to have, or should they be thought of as more objective values that must emerge from a structure of reason that may or may not be reflected in people's actual preferences?
- To what extent should prioritizing decisions be made within the framework of a general normative moral theory, or may they be made in more piecemeal fashion without an overarching moral framework?

Related to this last question, it might be noted that several moral and social theories have been referred to, used, or invoked in the previous description of various issues in the landscape of allocation decisions. Utilitarianism is the most obvious. Other theories are more “contractarian” and refer to the consent (perhaps the prior or hypothetical consent) of those who end up being governed by the policies in question. Still other approaches may be “intuitionist” in asserting basic claims about fairness or equality. It is not the intention of the author of this chapter to claim that any one of these frameworks is better or worse than another. They have been referred to only as a way of bringing to our explicit awareness certain underlying patterns of thought in allocation decisions.

References

- Brennan, Troyen A. (1991), *Just Doctoring: Medical Ethics in the Liberal State*, Berkeley: University of California Press.
- Broome, John (1982), Uncertainty in welfare economics and the value of life, in M. Jones-Lee (ed.), *The Value of Life and Safety*, proceedings of a conference held by the Geneva Association, Amsterdam: North-Holland, pp. 201–17.
- Core Services Committee of New Zealand (1994), *Core Services for 1995/1996*, Wellington: Ministry of Health.
- Culyer, A.J., Maynard, A., and Williams, A. (1981), Alternative systems of health care provision: an essay on motes and beams, in M. Olson (ed.), *A New Approach to the Economics of Health Care*, Washington, DC: American Enterprise Institute, pp. 131–50.
- Daniels, Norman (1985), *Just Health Care*, Cambridge: Cambridge University Press.
- Engelhardt, H. Tristram (1996), *The Foundations of Bioethics*, 2nd edn., New York: Oxford University Press.
- Hadorn, David (1991), Setting health care priorities in Oregon: cost-effectiveness meets the Rule of Rescue, *Journal of the American Medical Association* 265 (17): 2218–25.
- Hadorn, David (1992), The problem of discrimination in health care priority setting, *Journal of the American Medical Association* 268 (11): 1454–9.
- Hall, Mark A. (1997), *Making Medical Spending Decisions: The Law, Ethics, and Economics of Rationing Mechanisms*, New York: Oxford University Press.
- Harris, John (1985), *The Value of Life*, London: Routledge and Kegan Paul.
- Johannesson, J. and Johannesson, P.O. (1996), The economics of aging: on the attitude of Swedish people to the distribution of health care resources between the young and the old, *Health Policy* 37: 153–61.

- Jonsen, Albert (1986), Bentham in a box: technology assessment and health care allocation, *Law, Medicine and Health Care* 14: 172–4.
- Kilner, John (1990), *Who Lives? Who Dies? Ethical Criteria in Patient Selection*, New Haven, CT: Yale University Press.
- Menzel, Paul T. (1990), *Strong Medicine: The Ethical Rationing of Health Care*, New York: Oxford University Press.
- Menzel, Paul T. (2002), Justice and the basic structure of health-care systems, in R. Rhodes, M. Battin, and A. Silvers (eds.), *Medicine and Social Justice*, New York: Oxford University Press, pp. 24–37.
- Menzel, Paul T., Gold, M., Nord, E., Pinto-Prades, J.-L., Richardson, J., Ubel, P. (1999), Toward a broader view of cost-effectiveness analysis of health care, *Hastings Center Report* 29 (1): 7–15.
- Mishan, Ezra J. (1985), Consistency in the valuation of life: a wild goose chase? In E.F. Paul, J. Paul, and F.D. Miller, Jr. (eds.), *Ethics and Economics*, Oxford: Basil Blackwell, pp. 24–37.
- Morreim, E. Haavi. (1991), *Balancing Act: The New Medical Ethics of Medicine's New Economics*, Dordrecht: Kluwer.
- Nord, Erik (1993a), Health politicians do not wish to maximize health benefits, *Journal of the Norwegian Medical Association* 113: 1171–3.
- Nord, Erik (1993b), The relevance of health state after treatment in prioritising between different patients, *Journal of Medical Ethics* 19: 37–42.
- Nord, Erik (1999), *Cost–value Analysis in Health Care: Making Sense out of QALYs*, Cambridge: Cambridge University Press.
- Nord, Erik, Pinto-Prades, J.-L., Richardson, J., Menzel, P., Ubel, P. (1999), Incorporating societal concerns for fairness in numerical valuations of health programs, *Health Economics* 8: 25–39.
- Nord, Erik, Street, A., Richardson, J., Kuhse, H., Singer, P. (1996), The significance of age and duration of effect in social evaluation of health care, *Health Care Analysis* 4 (2): 103–11.
- Pellegrino, Edmund D., and Thomasma, David C. (1988), *For the Patient's Good: The Restoration of Beneficence in Health Care*, New York: Oxford University Press.
- Rawls, John (1971), *A Theory of Justice*, Cambridge, MA: Harvard University Press.
- Ubel, Peter A. (2000), *Pricing Life: Why It's Time for Health Care Rationing*, Cambridge, MA: MIT Press.
- Ubel, Peter, and Loewenstein, G. (1995), The efficacy and equity of retransplantation: an experimental survey of public attitudes, *Health Policy* 34: 145–51.
- Ubel, Peter, DeKay, M., Baron, J., and Asch, D. (1996), Cost-effectiveness analysis in a setting of budget constraints: is it equitable? *New England Journal of Medicine*, 334: 1174–7.
- Ubel, Peter, Scanlon, D., Loewenstein, G., and Kamlet, M. (1996), Individual utilities are inconsistent with rationing choices: a partial explanation of why Oregon's cost-effectiveness list failed, *Medical Decision making* 16 (2): 108–19.
- Williams, Alan (1985), Economics of coronary artery bypass grafting, *British Medical Journal* 291 (6491): 326–9.
- Williams, Alan (1997), Intergenerational equity: an exploration of the “fair innings” argument, *Health Economics* 6: 117–32.

Further Reading

- Daniels, Norman (1988), *Am I My Parents' Keeper? An Essay on Justice between the Young and the Old*, New York and Oxford: Oxford University Press.
- Eddy, David M. (1996), *Clinical Decision Making: From Theory to Practice*, Sudbury, MA: Jones and Bartlett.
- Gold, Marthe R., Siegel, J., Russell, L.B., and Weinstein, M. (eds.), *Cost-effectiveness in Health and Medicine*, New York: Oxford University Press.
- Kamm, Frances M. (1993), *Morality, Mortality*, vol. I, *Death and Whom to Save From It*, New York: Oxford University Press.