

CHAPTER 1.

CONCEPTS OF EFFICIENCY & EQUITY IN HEALTH ECONOMICS

1. Efficiency

Economics is the study of how individuals and societies choose to allocate scarce productive resources among competing alternative uses, and subsequently to distribute the 'products' from these uses among the members of a society. Health and health care are universally seen as two important products to which all societies commit resources. Health economics, therefore, is the study of how scarce resources are allocated among alternative uses for the care of sickness and the promotion, maintenance and improvement of health. It further includes the study of how health care and health-related services, their costs and benefits, and health itself are distributed among individuals and groups in society.

The term resources means the basic inputs to production – the time and abilities of individuals, land and natural resources (air, water, minerals, etc.), capital (facilities, equipment, etc.), and knowledge of production processes. Money, although it is an important medium of exchange and a very useful measuring device, is not defined by economists as a resource in itself. The importance of financial budgets, for example, stems from the command over resources which they confer upon those who control the budgets.

A fundamental problem facing all societies – and the reason that economics exists as an area of study – is that resources are scarce. Scarcity means that there are not, and can never be, enough resources to satisfy all human wants and needs. This observation is acutely clear every day when it comes to matters of illness and health, but it is equally true of other areas of human activity. There exists a constant conflict among alternative uses of resources, and a constant need to choose among alternative allocations.

Therefore, economists define the real cost of an activity (for example, provision of hospital services) in terms of the other outputs that must be given up (for example, other health services such as immunizations, or non-health services or commodities such as defence or vehicles) because resources are committed to it. Economists refer to this important basic concept as opportunity cost. Here are a couple of examples of opportunity cost. Policy makers in all societies face decisions about tradeoffs such as these on a daily basis:

Example 1. Opportunity Cost in the Allocation of Resources to Defence

The allocation of resources to the acquisition of new fighter aircraft for national defence in a developing country might preclude a range of other initiatives which could be to the nation's advantage. For example, a 50% reduction in the allocation of resources to new fighter aircraft could allow the achievement of a 15% decrease in the number of smokers in the country if programs modelled on recent successful programs elsewhere were introduced, or a 10% increase in the national literacy rate.

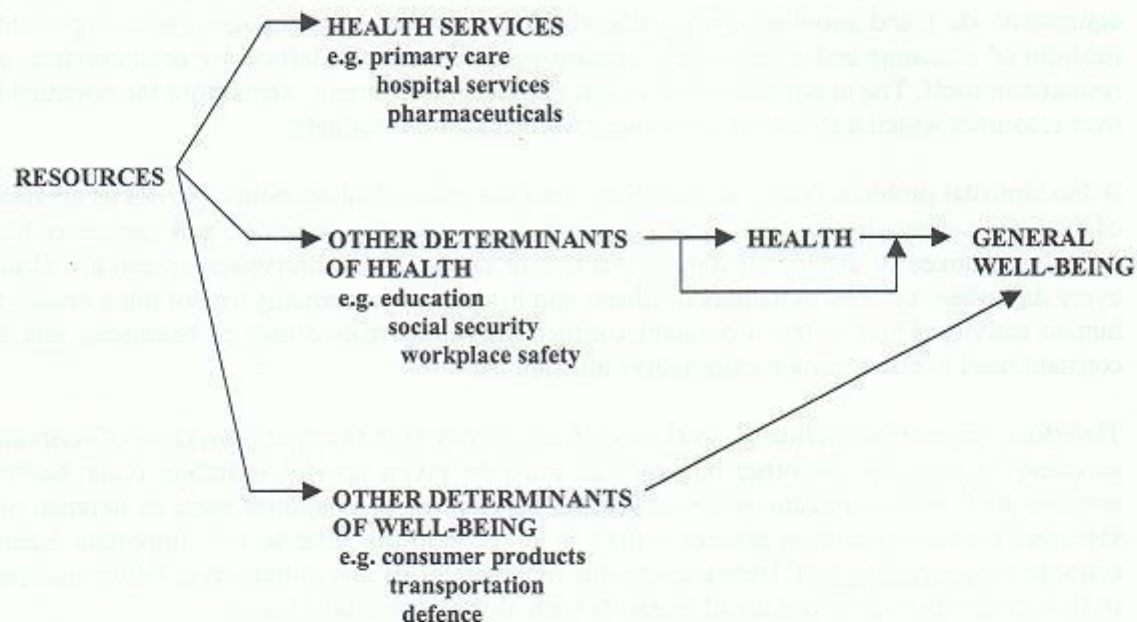
Example 2. Opportunity Cost in the Allocation of Resources to Different Types of Health Care

In the 1990-91 fiscal year, the provincial government of Ontario, Canada, granted a \$350 million budget increase for hospital services. A public health researcher pointed out that these funds could have been used to provide 70,000 publicly subsidized housing units for low-income families, or 547,000 publicly subsidized day care places for children, both of which can be viewed as alternative investments in health.

Two fundamental characteristics of economic analysis follow from the concept of opportunity cost. First, economics is concerned with evaluating and choosing among alternative courses of action, whether or not they are explicitly identified. Second, in doing so it examines both the costs and consequences of the alternatives.

The primary criterion that economics uses to organise and conduct these analyses is that of 'efficiency'. The basic concept of efficiency, as the word is understood in common usage by almost everyone, is quite simple: get the 'most' out of scarce resources! But beyond this intuitive advice it is not always clear what this involves, or how to achieve it, and – a word of warning at the outset – economists attach a very precise set of meanings to the concept of efficiency. Some of these meanings may not be obvious, agreed to or understood by everyone.

Before considering these meanings in more detail, it is worth pausing to reflect on the magnitude of the resource allocation problem facing societies once the implications of resource scarcity, competing uses for resources, and conflicting needs and wants are taken into account. This is illustrated with particular reference to the role of health and health care in the following diagram:



The diagram shows how resources can be used in alternative ways to 'produce' health care services, health and general well-being. The concept of efficiency and its meanings discussed below can be applied to each of these. Because efficiency is an instrumental concept, it is always necessary to specify clearly the outcome being sought or the 'output' being produced.

The diagram shows that resources can be used for a variety of purposes, divided here into three groups labelled 'health care services', 'other determinants of health', and 'other determinants of well-being'. Health care services contribute to general well-being through their effect on health, as do other determinants of health such as education, income security programs, and safe workplaces. These other determinants of health may also have a direct effect on general well-being. The third category of resource-using activities, other determinants of well-being, consists mainly of things that affect well-being directly, but some consumer products – e.g. food – can affect well-being indirectly by affecting health as well.

The three main elements of efficiency may be summarised in everyday language as follows:

1. Do not waste resources.
2. Produce each output in the least costly way.
3. Produce the types and amounts of output which people value most.

[Note: In the jargon of technical economics, we say that 1. involves being on rather than within an isoquant, 2. involves being at the point of tangency between an isoquant and an isocost line, and 3. involves being in a situation where marginal rates of transformation in production of outputs are equal to marginal rates of substitution in their consumption].

An efficient allocation of resources is one which simultaneously meets all three requirements. The first two requirements relate only to production. The third introduces consumption, thereby bringing together the demand and supply sides of the exchange of output. (The concepts of demand and supply are explained in more detail below).

The first element of efficiency above requires that for any given amount of output the amount of inputs used to produce it is minimised. (The requirement may also be stated such that maximum output is produced from any given combination of inputs). If this condition is not met, then it is possible either to obtain more output through a different configuration of resources, or to release some of the resources to alternative uses without sacrificing any current output. This element of efficiency is termed technical efficiency. Hospitals that are larger than they need to be to serve their communities are an example of technical inefficiency. The following case study provides another example in the context of the provision of health insurance in the US. In general, there will be several technically efficient combinations of inputs (e.g. combinations of labour and capital) for any given level of output:

Example 3. Technical Efficiency in the Provision of Health Insurance: The Case of the United States

Woolhandler and Himmelstein (1991) documented the technical inefficiency associated with the US system of health insurance, which relies on a multitude of private insurance companies to cover most Americans. When compared to a single-payer, publicly administered health insurance system like Canada's, several sources of technical inefficiency are apparent.

"...The existence of numerous insurers necessitates determinations of eligibility that would be superfluous if everyone were covered under a single, comprehensive program. Rather than a single claims-processing apparatus in each region, there are hundreds. Fragmentation also reduces the size of the insured group, limiting savings from economies of scale... Competition among insurers leads to marketing and cost-shifting, which benefit the individual insurance firm but raise systemwide costs...Co-payments, deductibles and exclusions are expensive to enforce..." (p. 1256-7)

"...A major advantage of public programs in terms of efficiency is their use of existing tax-collection structures, obviating the need for a redundant bureaucracy to collect money for health services." (p. 1257)

"...The existence of multiple payers in the US also imposes bureaucratic costs on health care providers. Hospitals must bill several insurance programs with varying and voluminous regulations on coverage, eligibility, and documentation. Moreover, billing on a per-patient basis requires an extensive internal accounting apparatus for attributing costs and charges to individual patients and insurers..." (p. 1257)

This technical inefficiency is very costly. Woolhandler and Himmelstein estimate that in the US administrative cost accounts for 19-24% of health care spending, while for Canada their estimate is 8-11%. If the US health insurance system was as technically efficient as Canada's, the annual savings today would be of the order of \$100 billion.

The second element of efficiency builds on the first but takes into account the relative cost of different inputs. It requires that inputs be combined so as to minimise the cost of any given output. (Alternatively, the requirement may be stated such that output is maximised for a given cost). For example, if labour is abundant and inexpensive relative to capital in one economy compared to another, then least-cost production methods will employ relatively more labour in the first economy. This element of efficiency can be termed cost-effectiveness. [Note: Economists do not always call this second element of efficiency 'cost-effectiveness'. They use the term 'technical efficiency' to include both this and the first concept described above]. For any given output in a particular setting, there will normally be only one combination of inputs that will be cost-effective. (It is only possible to claim that a specific combination of inputs is cost-effective in producing a particular output if it has been compared to one or more alternative combinations of inputs used for the same purpose). Note also that while cost-effectiveness can inform the question of how to produce an output at least cost, it does not address the question of whether or not the output should be produced. If something is not worth doing, it is not worth doing well!

The third element of efficiency links the supply of outputs to the demand for them by extending the analysis to consider the preferences and values of the members of society who consume the outputs. It requires that in addition to the achievement of technical efficiency and cost-effectiveness, resources be used to produce the types and amounts of outputs which best satisfy people i.e. which people value most highly. The term used by economists to describe this all-encompassing concept of efficiency is allocative efficiency (also often called Pareto efficiency). It is possible for an allocation of resources to be both technically efficient and cost-effective, but allocatively inefficient if producers are supplying too much or too little of any good or service relative to consumer's wishes. If mothers of young children want counselling services for behavioural problems instead of frequent well-child check-ups, then allocative efficiency might be improved by changing the mix of primary care services even if the well-child examinations were being provided cost-effectively.

In common language, then, efficiency means both 'doing things right' (technical efficiency and cost-effectiveness) and 'doing the right things' (allocative efficiency). The example below illustrates each of the three elements of efficiency in the context of primary health care.

Example 4. Efficiency in a Proposal for Primary Health Care Centres

Imagine you are a policy maker. While reviewing a proposal for the completion of a national network of primary health care centres, you become concerned about the efficiency of the centres, so you decide to use the three elements of efficiency (technical efficiency, cost-effectiveness, and allocative efficiency) to examine the proposal more closely.

First you look at staffing levels of physicians and nurses relative to studies of anticipated levels of communities' needs for primary care. If staffing levels were set too high relative to need, then some physicians or nurses might be idle for extended periods of the day, and this would be technically inefficient. (Technical inefficiency could also arise from the provision of inefficacious medical procedures, but you decide that you will need some help from medical experts to judge this aspect).

Staffing levels appear to be appropriate, so now you examine cost-effectiveness. You notice that the proposal plans for physicians to perform routine services like immunizations and well-child check-ups that nurses could perform with the same level of quality. But physicians are paid considerably more than nurses, so it looks like the cost-effectiveness of the proposal could be improved if the mix of providers was adjusted to utilise nurses more and physicians less.

Finally, you look at the types of services that will be provided in the centres, and you wonder about the allocative efficiency of the proposal. You know, for example, that some communities with existing centres have reported that mothers feel there are too few (sometimes no) counselling services about nutrition or

behavioural problems in children, even though they would prefer these services to some others that are currently provided. But counselling services are not on the list of services described in the proposal. Therefore, there may be room for improvement in allocative efficiency in the new (and existing) centres.

Moreover, to take health care services as an example, there is no single, correct, international solution for their efficient provision. Although all countries in theory have access to the same production knowledge, the relative cost of different inputs (e.g. nurses vs. doctors, drugs vs. hospital days) varies across countries, and citizen-consumers of different countries have different preferences and values. Therefore it is possible (indeed expected) that countries will display differences in how they provide services, in which services they provide, and to whom they are provided, independently of any differences in efficiency which may exist.

By necessity, statements about allocative efficiency involve value judgements about what criteria will be used to judge whether a particular resource allocation 'best satisfies' people, or is the 'most highly valued' or has 'too much or too little' of some goods and services. The standard criterion in economics comes from a branch of economic theory known as welfare economics. The criterion is known as the Pareto criterion, and states that allocative efficiency has been attained when it is not possible to change the allocation of resources to make any one person better off without making at least one other person worse off.

There are at least two other important characteristics of efficiency based on Paretian criteria. First, such a notion of efficiency is individualistic. Social welfare is assumed to be a function only of individual welfare, each individual is assumed to be the best judge of his or her gains and losses, and individual welfare is assumed to depend only on the goods and services the individual consumes. (In the real world, all of these assumptions are problematic. People care about the welfare of each other, their social groups, and their communities. Individuals often must make decisions on behalf of others. And even if considering only themselves, individuals care about other things – for example, the characteristics of the society they live in – not just the goods and services that they consume).

Second, the efficient outcomes realised under this concept of efficiency depend very much on the distribution of income and wealth among individuals in the society. In other words, under the Pareto criterion there is no unique allocation of resources that is the (one and only) efficient allocation. Rather, there is a set of efficient allocations, one for every different distribution of income and wealth. In practice, this difficulty can be avoided by accepting the existing distribution of income and wealth, although this is a very important value judgement that should be (but seldom is) made explicit in discussions of efficiency.

These characteristics of the standard economic approach to allocative efficiency mean that it may be possible for a society to prefer an 'inefficient' (in a Paretian sense) resource allocation to an 'efficient' one if, for example, the members of society judge it to be fairer in some way. A policy change that removed public subsidies for private hospitals in favour of expanding free public clinics might be an example of this. Another way to state this is that the choice among several allocatively efficient resource allocations must be made on the basis of criteria other than efficiency.

The usefulness of the Pareto criterion is very limited in practice because most changes in resource allocations do not in fact make some people worse off. That is, for almost any policy, there are both gainers and losers. In an attempt to extend the scope of the criterion, it has been proposed by some economists that a sufficient condition for viewing a change in resource allocation as an improvement in (allocative) efficiency is that those who gain from

the change value their gains enough to, in principle, be able to compensate the losers for the value of their losses, thereby leaving the losers as well off as before the change. This potential Pareto criterion does not require that the compensation actually be paid which many observers, including some economists, find ethically unacceptable. For example, suppose a policy of user fees improved access for the wealthy but reduced access for the poor. If the gains to the wealthy were sufficiently large to be able to compensate the poor, the potential Pareto criterion would deem this policy efficient even if the rich do not in fact compensate the poor for their reduced access. Nevertheless, it is this criterion which is the basis for economists' measurement technique of 'cost-benefit' analysis (described elsewhere in the course).

Economists are increasingly concerned about understanding the subjective and cultural underpinnings (rather than the consumption of goods and services by individuals) that societies use to judge whether changes in resource allocation are improvements. This is especially relevant in the area of health and health care. Considerations of fairness or equity are discussed below. Suffice here to point out that allocative efficiency as defined above does not necessarily imply social desirability, except under some very specific and controversial value judgements. Because value judgements and ethical principles are such an important part of criteria for allocative efficiency in the real world of policy makers, the measurement of equity is often inextricably involved in the measurement of efficiency. These measurement issues are addressed elsewhere in the course.

The general conditions for allocative efficiency under the potential Pareto criterion can be described in a technical way, using the economist's concept of the 'margin'. [Note: Economists define the marginal cost of an output to be the additional cost incurred in producing the last (or next) unit of that output. Similarly, the marginal benefit is the additional benefit obtained by consuming the last (or next) unit of an output. In an efficient world, marginal cost and marginal benefit are equal for each output, although they may vary across outputs. In other words, the value of an extra benefit that individuals and societies derive from the last unit of any output consumed is just equal to the opportunity cost of the resources (i.e. their value in their next best use) used up by producers to create that unit of output. For example, if a hospital wishes to expand its kidney dialysis program, consideration of allocative efficiency would require that it not expand past the point where the extra resources required (personnel, space, supplies and equipment) would create more benefit in another of the hospital's programs]. But they are intuitively quite simple and parallel (for society) a rule that individuals use every day in deciding how to allocate their time and money. Individually, we constantly make judgements about whether the extra benefit of doing something (e.g. buying another shirt, eating more dessert, visiting friends more often) is worth the extra cost involved. We tend to stop doing things when the extra costs exceed the extra benefits.

So too for societies pursuing their health care activities and health goals. For allocative efficiency, each activity or output should be pursued or produced only until the extra benefits from pursuing or producing it further are just equal to the extra costs.

Of course, this is a difficult determination to make in practice. An important contribution of economic theory has been the demonstration that this requirement can be satisfied (i.e. for each good or service produced, the marginal social cost equals the marginal social benefit), and a stable allocation of resources identified through a system of prices and markets. This 'solution' to the economic problems of what goods and services to produce (and in what

quantities), how to produce them, and how to distribute them applies only under some very specific conditions, however, which are described below.

When these conditions are violated, markets are said to 'fail' in that they do not lead to an allocatively efficient distribution of resources. Furthermore, unless compensatory policies are implemented, a system of prices and markets implicitly accepts the value judgements both that ability to pay and willingness to pay are appropriate criteria on which to base access to goods and services, and that the existing distribution of income and wealth, which facilitates purchases and consumption, is acceptable. In the case of health care, market failure is common, and the above value judgements are frequently rejected for certain categories of services. These issues are discussed further below.

2. Equity

Health economics is primarily concerned with finding the most efficient allocation of health resources to achieve a given policy goal (e.g. to improve population health, or protect individual health). However, efficiency is not the only criterion for judging the distribution of resources. Equity, or fairness, is a separate and important concern that health economists and other policy analysts must often address. Many jurisdictions (i.e. countries, states, provinces, districts) have formal mandates requiring that health care be distributed fairly or equitably (see e.g. Wagstaff and Van Doorslaer, 1993). However, decision makers do not always easily agree on what it means to be 'fair', either in principle or in practice. Policy questions concerning equity tend to be moral and ethical as well as technical in nature. The principles that health analysts use to ensure and assess equity derive from a variety of fields: philosophy, ethics, law, and political science, as well as economics.

Equity issues arise at many levels of the health care system. Different criteria, reasoning, and principles may be employed at the different levels. For example, a national ministry of health may be concerned with the fair allocation of budget dollars across regional health districts, as well as with an equitable distribution of professionals throughout the country. The regional health districts may be concerned with the fair allocation of primary and secondary health services between urban and rural populations, or among other groups such as women and children, the elderly, or the poor. Providers at local clinics may be concerned with the problem of triage: selecting individuals for priority treatment, for scarce drugs and other items and so forth. Each policy maker – public, administrative, and clinical – faces a unique set of defined health care goods, fair distribution processes and characteristics that identify needy or deserving recipients.

In this section we will survey basic concepts of equity and discuss how they might affect economic decision-making for health.

2.1. Analysing Equity

2.1.1. Basic elements of analysis

To distribute health (or health-producing) goods equitably means to distribute them according to acceptable processes or criteria. These processes and criteria relate to certain characteristics of the goods to be distributed, as well as to the characteristics of the recipients who will receive them. Policies for pursuing distributive equity address three basic analytic issues:

(1). Defining the *goods* to be distributed and characteristics that might affect their fair distribution.

An accepted method for distributing one type of good may not be acceptable for distributing another. Health care differs from other goods such as automobiles, food, entertainment, education and so forth. Various types of health care 'goods' also differ from each other, for example, health insurance, health care services, health outcomes. Relevant characteristics of goods include, for example, whether a good is seen as a basic necessity or a luxury, whether a good is very expensive or can be afforded by all, and so on.

(2). Identifying the appropriate *policy focus*.

A policy for ensuring equity may focus on one or both of two basic issues: (a). how to allocate (i.e. equitable processes) and (b). the acceptability of the resulting allocations (i.e. equitable end states). Equity objectives can be defined in terms of fair processes, fair end states, or both.

(3). Defining the relevant *characteristics of potential recipients* that might justify their claims to particular goods.

2.1.2. Types of goods

2.1.2.1. An overview of health care goods for distribution

For the purpose of equitable health sector policy planning, health care goods may be divided into various categories. The following are basic goods in health care whose distributions are often examined for equity:

(1). Health care insurance

(2). Health care inputs

- providers
- programs, services, interventions

(3). Access to health care insurance and health care

(4). Utilisation of health care

- use of services (or 'utilisation')
- use of effective services ('met needs')

(5). Benefits generated by (1), (2), (3), and (4)

- specific health effects ('outcomes')
- general benefits ('well-being')

In the production of health and well-being, each element listed depends to some extent on the previous ones: health effects depend on utilisation of the health system, utilisation depends upon individuals' and populations' access to services, access depends upon the existence of services. However, the goods are not exact proxies for each other i.e. access to health care insurance does not necessarily mean access to effective services, and achieving health benefits does not guarantee an improvement in well-being. Also, both health and well-being depend

upon resources besides health care – e.g. education, employment, nutrition, sanitation, social support, peace and so on.

Health policy makers ultimately seek a fair distribution of the 'highest' element on this list: an acceptable level of health and well-being shared by members of the community or country. Obviously, health and well-being are not 'things' that a government can distribute literally (that is, governments do not have a supply of health to hand out). However, governments often do have direct control over inputs (providers, programs, insurance) that can produce health and some indirect influence on individual behaviour (access to and use of services). To pursue equity in the health sector, therefore, policy makers strive to arrange services fairly and encourage appropriate access and utilisation – that is, they tend to focus on allocating the 'lower' goods of insurance, providers, access, etc. in the hope of influencing health and well-being. To monitor the success of a policy, however, policy makers ideally will examine not only the distribution of services and the rates of their use, but also the resulting distribution of the 'higher goods' of health and well-being.

2.1.2.2. Characteristics of goods that influence equitable distribution

The nature of a good will affect its equitable distribution. Important characteristics include the physical nature of the good, the degree to which a good has customised vs. generalised value across members of the community, and prevailing cultural beliefs about the good and its appropriate means of distribution. We can examine the equity implications of the health goods listed above with regard to each of these characteristics.

Physical nature of good: The physical nature of a good, in particular its divisibility and its scarcity, will affect how it can be distributed. For example, not only do policy makers not have a supply of health to distribute among the population, but genetic endowments mean that there are real physical limits to the extent to which any policies can equalise the distribution of health in a population. Among health services, the capital-intensive nature of hospitals makes it difficult (and non-sensical) to spread them evenly over a large geographic area. Hence, examining the distribution of hospitals themselves may give a distorted picture of access to hospital services, and policies to ensure access to hospital services must focus on goods needed to reach hospitals. Policy makers concerned with equitable access to hospitals in Zambia, for example, focus not only on the allocation of hospitals themselves, but also the allocation of other goods such as fuel and transportation needed to reach regionalised institutions.

In contrast, the workforce of health providers is a more divisible good. Although physicians and other health professionals tend to aggregate in urban centres, unlike hospitals, most are not bound there by physical necessity. Consequently, policy makers may realistically try to redistribute providers, particularly primary care providers, more equally over a geographic region.

Scarcity can affect policy makers' choices of fair allocation principles. In developing countries many high technologies such as kidney dialysis machines are scarce resources because their purchase and operating costs are high, and they require expertise to operate. Other resources are naturally scarce, whether expensive or not, for example, human organs and tissues (including blood). Where need outstrips the supply of scarce services such as dialysis, equity often focuses on fair processes for allocating the good. For example, attention

focuses on waiting lists, lotteries, or other methods for equitably allocating individuals' chances of dialysis (such methods are discussed below).

General vs. customised value: Goods may either have roughly the same value to everyone, or instead have customised value, where some would value the good more highly than others. In the health sector, access to health care is an important example of a good with generalised value. Policy makers ensuring equity try to distribute access such that there is equal access for all to at least basic services. In contrast, health care services themselves have customised value —the goal of policy makers is to ensure that only those who will benefit from the services receive them (i.e. not wasting them on those who would not benefit). Even for some health services, however, value may also accrue to others besides the recipient of the good, creating benefits or burdens on others besides the recipient. For example, an immunisation benefits the individual who receives it, but also other individuals who do not.

Cultural beliefs: Communities often have strong cultural beliefs that a particular good should be distributed in one manner, while another should be distributed in another. Some countries (such as Canada) value universal health insurance coverage as a symbol of national solidarity. Other countries (such as the US) feel a stronger ideological commitment to competitive markets than to universal coverage for health insurance. More generally, how would you react if you discovered that the manager of the district hospital's surgery waiting list was allowing individuals to buy and sell positions in the queue? Compare this to your reaction upon finding out that ushers at a local theatre were allowing people to buy and sell seat tickets to a hot-selling show? Most recoil at the former – it seems wrong and is prohibited in most settings – while the latter is much less troubling to most. Markets are seen as an equitable way to distribute some goods but not others, and equity considerations are one reason why we rely less on market mechanisms in the health sector than we do in most other sectors (below we will discuss some efficiency considerations as well). Even within the health care sector, policy makers make important distinctions between different types of health care services. For example, many governments are currently defining an 'essential' health care service package, which they would distribute broadly among the populace. These essential services are to be distinguished from non-essential or 'discretionary' services, which they believe are more legitimately distributed through the private market. Ideas about what services are 'essential' vary with the values and criteria of the policy makers. The World Bank, for example, has prescribed an essential package containing five types of service: prenatal and delivery care, family planning, management of the sick child, tuberculosis treatment, and management of sexually transmitted disease (World Bank, 1993).

Another illustration is the good of 'access to services': access barriers vary from community to community, ranging from cultural barriers such as language and education deficits to more material barriers such as transportation, funding, getting time off work etc. Distributing access more equally involves understanding the cultural norms about distributing these complementary goods.

2.1.3. Policy focus: fair processes vs. fair end-states

There are basically two possible policy focuses for ensuring fair resource allocation. The first focus is on developing fair *processes* for distributing goods. The second focus is on specifying fair *end-states* in the final distribution of goods. Health policy analysts can be concerned with either or both issues.

2.1.3.1. Process-principles for distributive equity

Process-principles of equity hold that fair procedures will necessarily distribute resources fairly. The belief is that as long as fair procedures are followed, the resulting distribution will, *de facto*, be fair. Process-principles for equity can be particularly important in two situations: (a). when the good being distributed cannot be divided to be shared among all those with a claim on it, so that a fair distribution focuses on providing each individual with a fair chance to obtain the good (leaving others to get none of it); (b). when informational problems preclude being able to make valid assessments of a good's final distribution, making end-state judgement impossible. The following example illustrates some commonly found process-principles for equity used in health care, at both the clinical and public policy levels:

In the country's largest teaching hospital, a cardiac surgeon performs a bypass surgery on a patient. When the patient develops complications, the surgeon puts the patient on a ventricular-assist pump. Once stabilised, she puts the patient on a waiting list for a heart transplant. The patient, who is wealthy, complains that the wait is too long and offers money to another patient ahead of her in the queue to take her spot. When this is rebuffed (by both the individual and the hospital administrators), she complains that she would rather take her chances in a lottery that would give each person waiting an equal chance of obtaining the transplant immediately. This is not possible, however, because not anybody can receive any heart. So she must wait her turn based on when her name was placed on the list and her clinical criteria.

Meanwhile, a public health official from a rural district of the province lobbies the health ministry to stop subsidising heart transplants in any manner (e.g. through hospital resources, training programs etc.). He insists that the money be redirected, among other things, to stemming an epidemic of malaria in his district. Furthermore, he calls for a process to be set up that would define essential services for which universal coverage will be provided. The Minister has advocated for a process in which policy makers would make the determination, while the public health official is advocating for a process that involves the general public.

This scenario brings up a number of issues. Among them is the process-principle of moral duty, or the imperative simply to 'do the right thing' in any given situation, a principle upon which many health professionals place great emphasis. In health care, moral duties are commonly part of professional role expectations. The cardiac surgeon feels a moral obligation not to abandon the patient in her care. The public health official feels a moral obligation to advocate for improving the health of his district population. In particular, clinicians' ethical duties to rescue individuals in danger can redirect large amounts of resources into intensive or emergency care. To the extent that the community recognises this particular moral duty as fair and just, it may accept the resulting reallocations as equitable. However, as a guide to resource allocation, moral duties often conflict with other criteria for equity as well as efficiency.

This example raises the process-principle of a lottery, such as that requested by the heart patient. Though seldom used in clinical practice (as opposed to clinical research, where it underlies randomised control trials), lotteries are seen by some as a fair way to distribute an indivisible good because they give each candidate an equal chance of receiving a scarce good, rather than an equal amount of the good itself. An advantage of the lottery is its blindness to the characteristics of recipients – everyone has the same probability of winning regardless of who they are. Consequently, there is little chance of unfair discrimination. Lotteries are not used frequently, in part because people do not always want an equal chance – they want a fair chance that reflects what they see as their legitimate claim to the good. This relates closely to an issue we explore below, where health policy makers concerned with equitable end-states (as opposed to processes) often seek *fair discrimination* in the allocation of health care among candidates, which lotteries also prevent.

Queuing is an alternative process for allocating a scarce and needed resource that reflects these considerations. Queuing in health care is often not based solely on length of time on the waiting list, but is mediated by special characteristics of the recipients, such as the severity of their condition (again, these discriminating characteristics are discussed below).

Another process-principle for equity is free exchange, ideally conducted through a perfect market. This issue arises frequently in health policy, especially with regard to the allocation of health insurance and services among the wealthy. Market proponents argue that such exchanges do allocate resources equitably if: (1). Transactions are completely voluntary; (2). People have full information about the consequences of their transactions; (3). There are no substantial starting differences in people's wealth. These conditions are rarely satisfied in markets for any health care good. The philosophical underpinnings of free-exchange processes, and the problems in applying them to necessarily-imperfect health care markets are discussed below.

Governance and decision making processes, such as that related to defining services to be covered, may be structured to ensure equitable resource planning. Various processes may be used. For example, democratic processes may be used to involve the general public in broad policy mandates. In health care rationing, contractarian processes have become a popular ideal for making fair rationing decisions (Rawls, 1971). Such processes require decision makers to specify an allocation of goods as if they had no idea what situation they might find themselves in after the goods have been allocated. Simply put, decision makers must choose as if 'there but for fortune go I'. This approach tends to some extent to minimise harm to whoever will be made worse off. For example, a limited service package for the poor in the US state of Oregon has been criticised on the grounds that the poor were not adequately involved in the rationing decisions, and that the policy makers (state employees) might not agree to such a limited package for their own health benefits (through state employee health insurance). (See e.g. Daniels, 1991).

2.1.3.2. End-state principles for distributive equity

When policy makers focus on end-state aspects of distributive equity, they are less concerned with the process of distribution and more concerned with the end-state distribution, that is, where the resources eventually end up. Important end-state principles can be illustrated by considering the problem of allocating a central budget for health care. In many countries, central health care budgets are being devolved to the districts. What is the fairest way to divide the provincial budget among the districts? Although equal per capita grants are sometimes used, analysts have more recently focused on adjusting per capita grants to reflect variations in both the relative health care needs across districts and in the cost of providing services. Zambia, for instance, adjusts its per capita grants for variation across districts in population density, fuel prices, the availability of a bank, and the cholera proneness of the district. In the province of British Columbia, the regional funding formula adjusts for the age-sex composition of the population and for variation in population health need as measured by standardized mortality ratios. And with the UK National Health Service, a portion of the regional budget allocation has for years been based in part on standardized mortality ratios as an indicator of morbidity (Mays, 1989).

Equality is a key concept for analysing end-state distributions. Nearly all end-state distribution principles emphasise the importance of equalising the distribution of some goods

among certain recipients (Sen, 1992). Rival theories dispute what the good is. Hence, although they are closely related, equality is not the same thing as equity. A strictly equal distribution of a budget may seem quite unfair. For instance, many districts could argue that their situation is special with regard to some key characteristic, and so they have a legitimate claim to more resources for health. Providing unequal per capita allocations that correlate with needs and costs, for example, may be necessary in order to provide equal access for residents of each district.

Aristotle first articulated the idea of "formal justice", which defines an equitable end-state distribution as one in which equals are treated equally, and unequals are treated unequally according to their degree of inequality. In economics, the term horizontal equity is often used to refer to the distribution of equal amounts of a good among recipients who are similarly situated according to relevant aspects of their situations. Hence, the funding formula attempts to ensure that those with similar needs and costs receive similar per capita amounts. In contrast, vertical equity refers to the imperative to distribute unequal amounts among differently-situated recipients in proportion to the degree that they are differently situated. Hence, the formula similarly attempts to ensure that those districts with greater health care needs receive a greater share of the budget.

The following example discusses how some of these process and end-state principles are embodied in a recent WHO policy document relating to the allocation of scarce organs:

Example 5. Organ Distribution Policy: A Focus on Equitable Processes

Organs are scarce everywhere relative to the number of people who could benefit from a transplant, consequently organ allocation policies offer excellent case studies in how equity principles can be applied. The following excerpts from policies recommended by the World Health Organisation (WHO, 1991) illustrate both processes and end-states that make organ allocation policies equitable.

"The human body and its parts cannot be the subject of commercial transactions. Accordingly, giving or receiving payment (including any other compensation or reward) for organs should be prohibited." (p.1470).

This passage points to unacceptable processes for exchange i.e. markets cannot be used for the distribution of organs, and it is not acceptable to exchange organs for other goods such as cash or other rewards.

"An organ may be removed from the body of an adult living donor for the purpose of transplantation if the donor gives free consent. The donor should be free of any undue influence and pressure and sufficiently informed to be able to understand and weight the risks, benefits, and consequences of consent." (p. 1470).

"It should be prohibited for physicians and other health professionals to engage in organ transplantation procedures if they have reason to believe that the organs concerned have been the subject of commercial transactions." (p. 1471).

These passages invoke moral duty principles. The organ donation process is fair if the donor is free to decide voluntarily and is adequately informed. Providers are obliged not to use organs that have been procured illegitimately.

"In the light of the principles of distributive justice and equity, donated organs should be made available to patients on the basis of medical need and not on the basis of financial or other considerations." (p. 1471).

This passage focuses more on fair end-states, and in particular, the appropriate characteristics of organ recipients. People with similar medical needs should have a similar chance of getting an organ. Financial characteristics, such as patients' incomes, should not affect their chances of getting an organ.

How would one judge the horizontal and vertical equity of these different systems? Obviously, to apply the concepts of vertical and horizontal equity, analysts must define what aspects of potential recipients' situations are relevant to equity judgements and what aspects are not. The above discussion of funding formulae focused on need and costs, but the relevant aspects vary across settings. This leads us to our final problem for analysing equity: identifying the appropriate characteristics of potential recipients.

2.1.4. Characteristics of potential recipients

Which recipient characteristics should legitimate their claims to a greater portion of health resources? Aiming at an equitable end-state distribution requires discrimination among recipients. "Recipients" may be thought of either as individuals, or, more commonly for the purpose of public policy making, as populations or groups. Whether a particular basis for discrimination (e.g. level of need, geographic location, gender, age, income level) is fair or 'unfair' will depend upon the communities' commitments to the groups or individuals in question. Consider the following situation:

AIDS has decimated many communities, afflicting the poor, working, and professional classes alike. A provincial ad hoc committee is reviewing how to allocate scarce anti-viral AIDS drugs. During a community consultation, one district representative requests that the province give health care workers priority access to the drugs. She argues that this is only fair because health professionals work the hardest to alleviate the AIDS epidemic. Another presenter argues that all professional and managerial workers should get preferential access. He argues that the country has invested disproportionately in their education and training and that their early deaths are creating a national 'brain drain'.

In this example, the characteristic at issue is the individual's contribution to society, and the equity principle is that health care should be given in proportion to an individual's contribution, either past, present, or future. Policy analysts have noted a number of moral pitfalls associated with assessing 'contributions' using human capital criteria to distribute health care and other social goods (Robinson, 1986). A major problem is that health and social productivity are interrelated: sick individuals have a harder time contributing, so preferential access for the 'productive' also often means preferential access for the 'healthy'. Another problem is that broader socioeconomic conditions – including historical discrimination – can make certain classes of people (e.g. women, ethnic minorities) less productive in measurable terms (e.g. taxable earnings, intellectual contributions). The contribution principle is reflected, for example, in the recommendation that days of healthy life generated by a health care intervention be weighted by the age-specific wage rate of those who are affected. However, rightly or wrongly, contribution principles often implicitly guide the distribution of care, and analysts should be able to recognise this commonly argued rationale for preferential access.

Other important characteristics can be illustrated through the following scenario:

A rural district is developing a central plan for distributing primary health clinics. In one small village, there currently exist two health clinics, each of which caters for a particular religious group. Due to one religion's prohibitions against casual association between men and women, its own clinic is further divided between the two genders: one group of clinicians and rooms is reserved for women and children, the other for men. District analysts have determined that closing one clinic, and requiring the remaining

clinic to serve all village members, would improve efficiency and quality. Activists from each religious community have protested that their own clinic must remain open and exclusive.

This example illustrates the characteristics of group membership. In equity analysis, group-based end-state distributions hold that people across groups should be treated alike, or alternatively, that certain groups should have preferential access to certain types of resources. Groups may be defined socially (e.g. by community, religious, or other group), economically (e.g. wealthy or impoverished), demographically (e.g. by gender, age, race, urban/rural), or along other dimensions. In Tanzania, for example, an analysis that divided the population by income level found that a little over a third of the users of government funded hospitals are wealthy citizens who would be able to pay for hospital care themselves. Analysts criticise the government subsidisation of such services as inequitable, arguing that government subsidies should be reserved for the poor.

The groups deemed relevant to an equity analysis in a particular context will be dictated by factors such as the good being distributed, cultural norms, ideologies, and beliefs about historical entitlements or injustices. The meaning and legitimacy of a particular group division is best understood within the particular social context. Group-based equity analysis resource allocation obviously can be used for social justice or prejudicial discrimination, and applications of certain characteristics can conflict with international human rights principles.

The discriminating characteristic most frequently used to define equitable distributions of health care is need. Although the precise meaning of "need" is debated, its relevance to health care seldom is. In extreme situations where medical personnel and supplies are scarce (e.g. in a war zone), to allocate these resources fairly medics 'triage' the injured on the basis of the severity of their injuries and the likelihood that they will benefit from emergency medical attention. And this same notion can be used more generally: need is most often understood in terms of an individual's or population's ability to benefit from health care – if the use of a health care service can be expected to improve that individual's health or well-being, then it is said to be needed. Of course, the health care needs of an individual over a lifetime can be enormous and constrained only by the health care system's technical capabilities and effectiveness. Policy makers allocating limited resources inevitably must decide whose needs will be met and whose will not.

Need may be conceived in different ways, depending upon the good to be distributed. For example, if the 'good' is the financial subsidisation of services, then neediness may be defined in part by degree of poverty. In the case of health care resources themselves, to determine whether health services are fairly distributed, one would ideally like to divide the population by degree of health care need, but such information is seldom directly available. Therefore, analyses of equity commonly compare distributions of goods between wealthy vs. poor, old vs. young, one ethnic group vs. another, urban vs. rural, male vs. female, and so on, each of which may serve as proxies for certain aspects of need. Each of these groups may be characterised by particular health or more general resource needs, indicating that they could benefit from a particular portfolio of health services.

Remember that equitable end-state distributions would allocate the same amount of resources to people in similar situations (horizontal equity), but different amounts of resources to people in different situations, according to their difference in need (vertical equity). The distribution of vaccination coverage, for example, has been used as a measure of inequity in health services between urban and rural populations in Peru (Musgrove, 1986). Residents of urban regions are vaccinated at approximately four times the rate of rural residents for poliomyelitis

and other diseases. Because the biological need for vaccination varies little among individuals or groups, analysts can assume that members of rural and urban regions need vaccination equally. The large variations in vaccination rates, then, reveal inequitable provision of health care between urban and rural populations.

Suppose, however, that asthma drug dispensing also varied fourfold between urban and rural populations. Because need is the appropriate discriminating characteristic, it is less clear that this particular inequality suggests inequity. Analysts would need more information about the epidemiology of asthma in the two populations: if the poor quality of urban air causes four times the prevalence of asthma in urban regions, then the distribution of asthma drugs may be quite equitable.

A thorough analysis of the equity of any health care system would involve segmenting the populations (as well as the goods) in a number of different ways. There is no single, agreed-upon dimension along which to divide populations with regard to either their demographic, socioeconomic, or even health characteristics. The most sensible division will depend upon the values and commitments of the government doing the analysis as well as the particular good in question.

Bibliography and References for Chapter 1 (Optional)

Daniels, N., 1991, "Is the Oregon rationing plan fair?", JAMA 265: 2232-2235.

Mays, N., 1989, "Measuring Morbidity for Resource Allocation", BMJ 295: 703-706.

Musgrove, P., 1986, "Measurement of Equity for Health", World Health Statistics Quarterly 39: 325-335.

Rawls, J., 1971, A Theory of Justice. (Cambridge, Massachusetts: Harvard University Press).

Robinson, J., 1986, "Philosophical Origins of the Economic Valuation of Life", Millbank Quarterly 64: 133-155.

Sen, A., 1992, Inequality Re-examined. (Cambridge MA: Harvard University Press).

Wagstaff, A., and Van Doorslaer, E., 1993, "Equity in the Finance and Delivery of Health Care: Concepts and Definitions", in Van Doorslaer, E., Wagstaff, A., and Rutten, F., (eds.) Equity in the Finance and Delivery of Health Care: An International Perspective. (Oxford: Oxford University Press).

Woolhandler, S., and Himmelstein, D., 1991, "The Deteriorating Efficiency of the US Health Care System", New England Journal of Medicine 324: 1253-1258.

World Bank, 1993, "Clinical Services", in Investing in Health: World Development Indicators. (New York: Oxford University Press, p. 108-133).

World Health Organisation, 1991, "Guiding Principles on Human Organ Transplantation", The Lancet 337: 1470-1471.