

## CHAPTER TWO

# *How Well do Health Systems Perform ?*

*Better health is unquestionably the primary goal of a health system. But because health care can be catastrophically costly and the need for it unpredictable, mechanisms for sharing risk and providing financial protection are important. A second goal of health systems is therefore fairness in financial contribution. A third goal – responsiveness to people’s expectations in regard to non-health matters – reflects the importance of respecting people’s dignity, autonomy and the confidentiality of information. WHO has engaged in a major exercise to obtain and analyse data in order to assess how far health systems in WHO Member States are achieving these goals for which they should be accountable, and how efficiently they are using their resources in doing so. By focusing on a few universal functions that health systems undertake, this report provides an evidence base to assist policy-makers improve health system performance.*

## 2

## HOW WELL DO HEALTH SYSTEMS PERFORM?

### ATTAINMENT AND PERFORMANCE

Assessing how well a health system does its job requires dealing with two large questions. The first is how to measure the outcomes of interest – that is, to determine what is achieved with respect to the three objectives of good health, responsiveness and fair financial contribution (*attainment*). The second is how to compare those attainments with what the system *should* be able to accomplish – that is, the best that could be achieved with the same resources (*performance*). Although progress is feasible against many of society’s health problems, some of the causes lie completely outside even a broad notion of what health systems are. Health systems cannot be held responsible for influences such as the distribution of income and wealth, any more than for the impact of the climate. But avoidable deaths and illness from childbirth, measles, malaria or tobacco consumption can properly be laid at their door. A fair judgement of how much health damage it should be possible to avoid requires an estimate of the best that can be expected, and of the least that can be demanded, of a system. The same is true of progress towards the other two objectives, although much less is known about them (1).

### GOALS AND FUNCTIONS

Better health is of course the *raison d’être* of a health system, and unquestionably its primary or defining goal: if health systems did nothing to protect or improve health there would be no reason for them. Other systems in society may contribute greatly to the population’s health, but not as their primary goal. For example, the education system makes a large difference to health, but its defining goal is to educate. Influence also flows the other way: better health makes children better able to learn, but that is not the defining purpose of the health system. In contrast, the goal of fair financing is common to all societal systems. This is obvious when the system is paid for socially, but it holds even when everything is financed purely by individual purchases. It is only the notion of fairness that may vary. “Getting what you pay for” is generally accepted as fair in market transactions, but seems much less fair where health services are concerned. Similarly, in any system, people have expectations which society regards as legitimate, as to how they should be treated, both physically and psychologically. Responsiveness is therefore always a social goal. Taking the education system as an example, fair financing means the right balance of contributions from households which do and those which do not have children in school, and enough subsidy that poor children are not denied schooling for financial reasons. Responsiveness includes respect for parents’ wishes for their children, and avoiding abuse or humiliation of the students themselves.

The health system differs from other social systems such as education, and from the markets for most consumer goods and services, in two ways which make the goals of fair financing and responsiveness particularly significant. One is that health care can be catastrophically costly. Much of the need for care is unpredictable, so it is vital for people to be protected from having to choose between financial ruin and loss of health. Mechanisms for sharing risk and providing financial protection are more important even than in other cases where people buy insurance, as for physical assets like homes or vehicles, or against the financial risk to the family of a breadwinner dying young. The other peculiarity of health is that illness itself, and medical care as well, can threaten people's dignity and their ability to control what happens to them more than most other events to which they are exposed. Among other things, responsiveness means reducing the damage to one's dignity and autonomy, and the fear and shame that sickness often brings with it.

Systems are often charged to be affordable, equitable, accessible, sustainable, of good quality, and perhaps to have many other virtues as well. However, desiderata such as accessibility are really a means to an end; they are instrumental rather than final goals. The more accessible a system is, the more people should utilize it to improve their health. In contrast, the goals of health, fair financing, and responsiveness are each intrinsically valuable. Raising the achievement of any goal or combination of goals, without lowering the attainment of another, represents an improvement. So if the achievement of these goals can be measured, then instrumental goals such as accessibility become unnecessary as proxy measures of overall performance; they are relevant rather as explanations of good or bad outcomes.

It is certainly true that financing that is more fairly distributed may contribute to better health, by reducing the risk that people who need care do not get it because it would cost too much, or that paying for health care leaves them impoverished and exposed to more health problems. And a system that is more responsive to what people want and expect can also make for better health, because potential patients are more likely to utilize care if they anticipate being treated well. Both goals therefore are partly instrumental, in that they promote improvements in health status. But they would be valuable even if that did not happen. That is, paying equitably for the system is a good thing in itself. So is assuring that people are treated promptly, with respect for their dignity and their wishes, and that patients receive adequate physical and affective support while undergoing treatment. The three goals are separable, as is often shown by people's unhappiness with a system even when the health outcomes are satisfactory.

Comparing how health systems perform means looking at what they achieve and at what they *do* – how they carry out certain *functions* – in order to achieve anything (2). These functions could be classified and related to system objectives in many different ways. For example, the “Public health in the Americas” initiative led by the Pan American Health Organization describes 12 different “essential functions”, and proposes between three and six sub-functions for each one (3). Many of these functions correspond to the task of stewardship which this report emphasizes, others to service provision and to resource generation. The four functions described in this chapter embrace these and other more specific activities. Figure 2.1 indicates how these functions – delivering personal and non-personal health services; raising, pooling and allocating the revenues to purchase those services; investing in people, buildings and equipment; and acting as the overall stewards of the resources, powers and expectations entrusted to them – are related to one another and to the objectives of the system. Stewardship occupies a special place because it involves oversight of all the other functions, and has direct or indirect effects on all the outcomes. Comparing the way these functions are actually carried out provides a basis for understanding

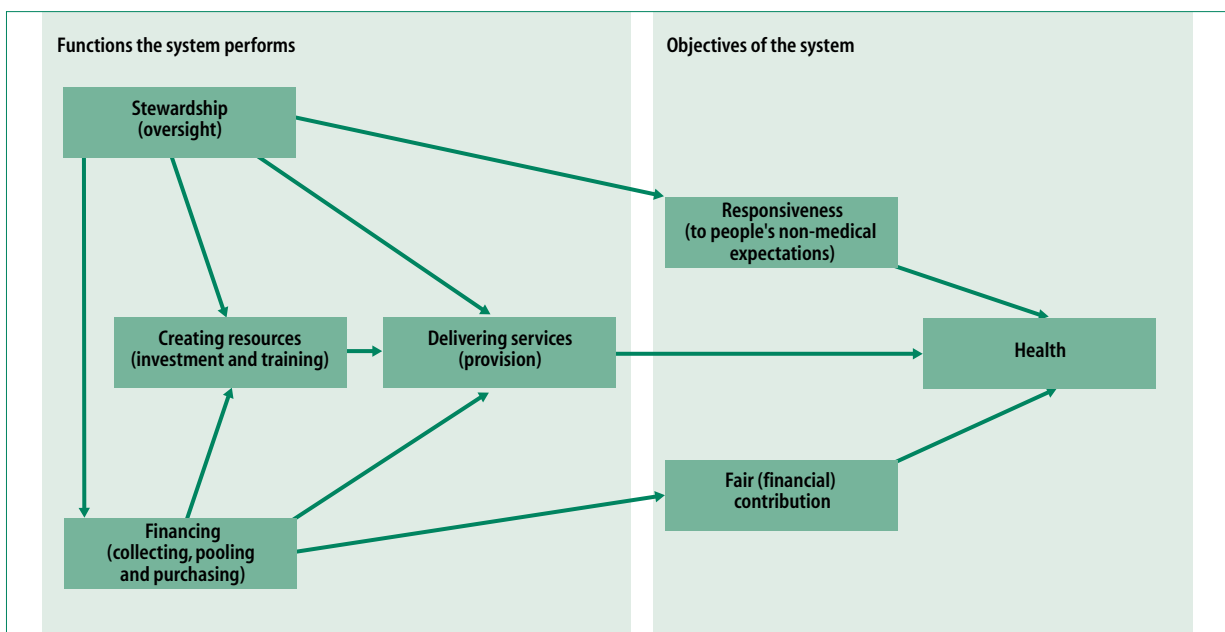
performance variations over time and among countries. Some evidence about these functions, and how they influence the attainment of the fundamental objectives in different health systems, is examined in the next four chapters.

In the view of most people, the health system is simply those providers and organizations which deliver personal medical services. Defining the health system more broadly means that the people and organizations which deliver medical care are not the whole system; rather, they exercise one of the principal functions of the system. They also share, sometimes appropriately and sometimes less so, in the other functions of financing, investment and stewardship. The question of who should undertake which functions is one of the crucial issues treated in later chapters.

It is common to describe the struggle for good health in quasi-military terms, to talk of “fighting” malaria or AIDS, to refer to a “campaign” of immunization or the “conquest” of smallpox, to “free” a population or a geographical area of some disease, to worry about the “arms race” that constantly occurs between pathogens and the drugs to hold them in check, to hope for a “silver bullet” against cancer or diabetes. In those terms, the providers of direct health services – whether aimed at individuals, communities or the environment – can be considered the front-line troops defending society against illness. But just as with an army, the health system must be much more than the soldiers in the field if it is to win any battles. Behind them is an entire apparatus to ensure that the fighters are adequately trained, informed, financed, supplied, inspired and led. It is also crucial to treat decently the population they are supposed to protect, to teach the “civilians” in the struggle how to defend themselves and their families, and to share equitably the burden of financing the war.

Unless those functions are properly carried out, firepower will be much less effective than it might be, and casualties will be higher. The emphasis here on overall results and on the functions more distant from the front line does not mean any denigration of the importance of disease control. It means rather to step back and consider what it is that the system as a whole is trying to do, and how well it is succeeding. Success means, among other things, more effective control of diseases, through better performance.

**Figure 2.1 Relations between functions and objectives of a health system**



## GOODNESS AND FAIRNESS: BOTH LEVEL AND DISTRIBUTION MATTER

A good health system, above all, contributes to good health. But it is not always satisfactory to protect or improve the average health of the population, if at the same time inequality worsens or remains high because the gain accrues disproportionately to those already enjoying better health. The health system also has the responsibility to try to reduce inequalities by preferentially improving the health of the worse-off, wherever these inequalities are caused by conditions amenable to intervention. The objective of good health is really twofold: the best attainable average level – *goodness* – and the smallest feasible differences among individuals and groups – *fairness*. A gain in either one of these, with no change in the other, constitutes an improvement, but the two may be in conflict. The logic is somewhat parallel to that concerning the distribution of income in a population. It is desirable to raise the average level, to reduce inequality, or both, and sometimes to judge the relative values of one and the other goal (with the difference that there is no argument for taking health away from anyone – health, unlike income or nonhuman assets, cannot be directly redistributed).

The distinction between the overall level and how it is distributed in the population also applies to responsiveness. Goodness means the system responds well on average to what people expect of it, with respect to its non-health aspects. Fairness means that it responds equally well to everyone, without discrimination or differences in how people are treated. The distribution of responsiveness matters, just as the distribution of health does. Either one is valuable by itself.

In contrast to the objectives of good health and responsiveness, there is no overall notion of goodness related to financing. There are good and bad ways to raise the resources for a health system, of course, but they are more or less good primarily as they affect how fairly the financial burden is shared. Fair financing, as the name suggests, is concerned only with distribution. It is not related to the total resource bill, nor to how the funds are used. While it is unambiguously preferable to have better health or a higher level of responsiveness, it is not always better to spend more on health because at high levels of expenditure there may be little additional health gain from more resources. The objectives of the health system do not include any particular level of total spending, either absolutely or relative to income. This is because, at all levels of spending, the resources devoted to health have competing uses, and it is a social choice – with no correct answer – how much to allocate to the health system. Nonetheless there is probably a minimum level of expenditure required to provide a whole population with a handful of the most cost-effective services, and many poor countries are currently spending too little even to assure that (4).

In countries where most health financing is private, and is largely out of pocket, no one makes this choice overall; it results from millions of individual decisions. As the level of prepayment rises, there are fewer and larger decisions, because spending is more and more determined by the policies and budgets of public entities and insurance funds. The public budget decision has the greatest effect in high income countries where most funding is government controlled or mandated, but in all countries it is one of the most basic public decisions. It is something that can be directly chosen, as the level of health outcome or of responsiveness cannot be.

## MEASURING GOAL ACHIEVEMENT

To assess a health system, one must measure five things: the overall level of health; the distribution of health in the population; the overall level of responsiveness; the distribution of responsiveness; and the distribution of financial contribution. For each one, WHO has used existing sources or newly generated data to calculate measures of attainment for the countries where information could be obtained. These data were also used to estimate values when particular numbers were judged unreliable, and to estimate attainment and performance for all other Member States. Several of these measures are novel and are explained in detail in the Statistical Annex, where all the estimates are given, along with intervals expressing the uncertainty or degree of confidence in the point estimate. The correct value for any indicator is estimated to have an 80% probability of falling within the uncertainty interval, with chances of 10% each of falling below the low value or above the high one. This recognition of inexactness underscores the importance of getting more and better data on all the basic indicators of population health, responsiveness and fairness in financial contribution, a task which forms part of WHO's continuing programme of work.

The achievements with respect to each objective are used to rank countries, as are the overall measures of achievement and performance described below. Since a given country or health system may have very different ranks on different attainments, Annex Table 1 shows the complete ranking for all Member States on all the measures. In several subsequent tables, countries are ranked in order of achievement or performance, and the order varies from one table to another. Since the ranking is based on estimates which include uncertainty as to the exact values, the rank assigned also includes uncertainty: a health system is not always assigned a specific position relative to all others but is estimated to lie somewhere within a narrower or broader range, depending on the uncertainties in the calculation. The ranks of different health systems therefore sometimes overlap to a greater or lesser degree, and two or more countries may have the same rank.

*Health* is the defining objective for the health system. This means making the health status of the entire population as good as possible over people's whole life cycle, taking account of both premature mortality and disability. Annex Table 2 presents three conventional and partial measures of health status, by country, without ranking: these are the probability of dying before age five years or between ages 15 and 59 years, and life expectancy at birth. For the first time, these measures are presented with estimates of uncertainty, and these uncertainties carry over to subsequent calculations. On the basis of the mortality figures, five strata are identified, ranging from low child and adult mortality to high child mortality and very high adult death rates. Combining these strata with the six WHO Regions gives 14 subregions defined geographically and epidemiologically (see the list of Member States by WHO Region and mortality stratum). Annex Table 3 presents estimates of mortality by cause and sex in 1999 in each of these subregions (not by country), and Annex Table 4 combines these death rates with information about disability to create estimates of one measure of overall population health: the burden of disease, that is, the numbers of disability-adjusted life years (DALYs) lost.

To assess overall population health and thus to judge how well the objective of good health is being achieved, WHO has chosen to use disability-adjusted life expectancy (DALE), which has the advantage of being directly comparable to life expectancy estimated from mortality alone and is readily compared across populations. Annex Table 5 provides estimates for all countries of disability-adjusted life expectancy. DALE is estimated to equal or exceed 70 years in 24 countries, and 60 years in over half the Member States of WHO. At



the other extreme are 32 countries where disability-adjusted life expectancy is estimated to be less than 40 years. Many of these are countries with major epidemics of HIV/AIDS, among other causes. Box 2.1 describes how these summary measures of population health are constructed and how they are related.

Figure 2.2 summarizes the relation between DALE and life expectancy without adjustment, for each of the 14 subregions, for both men and women. The adjustment is nearly uniform, at about seven years of healthy life equivalent lost to disability. Both absolutely and relatively this loss is slightly less for richer, low-mortality subregions, despite the fact that people live longer there and so have more opportunity to acquire non-fatal disabilities. Disability makes a substantial difference in poorer countries because some limitations – injury, blindness, paralysis and the debilitating effects of several tropical diseases such as malaria and shistosomiasis – strike children and young adults. Separating life expectancy into years in good health and years lived with disability therefore widens rather than narrows the difference in health status between richer and poorer populations. This is most evident in the share of life expectancy which is lost to disability: it ranges from less than 9% in the healthiest subregions to more than 14% in the least healthy. Annex Table 5 shows these shares for individual countries, where the range is even wider.

Annex Table 5 also provides estimates of health inequality. The distributional measure of health ranges from 1 for the case of perfect equality to zero for extreme inequality, which corresponds to a fraction of the population having an expectancy of 100 years and the rest

### Box 2.1 Summary measures of population health

No measure is perfect for the purpose of summing up the health of a population; each way of estimating it violates one or another desirable criterion. The two principal approaches are the burden of disease, which measures losses of good health compared to a long life free of disability, and some measure of life expectancy, adjusted to take account of time lived with a disability. Both ways of summarizing health use the same information about mortality and disability, and both are related to a survivorship curve, such as the bold line between the areas labelled *Disability* and *Mortality* in the figure.

The area labelled *Mortality* represents losses due to death, compared to a high standard of life expectancy: the burden of disease corresponds to all of that area plus a fraction of the area corresponding to time lived with disability. The fraction depends on the disability weights assigned to various states between death and perfect health. Life expectancy without any adjustment corresponds to the areas labelled *Survival free of disability* and *Disability* together, the whole area under the survivorship curve. Disability-adjusted life expectancy (DALE) then corresponds to the area for sur-

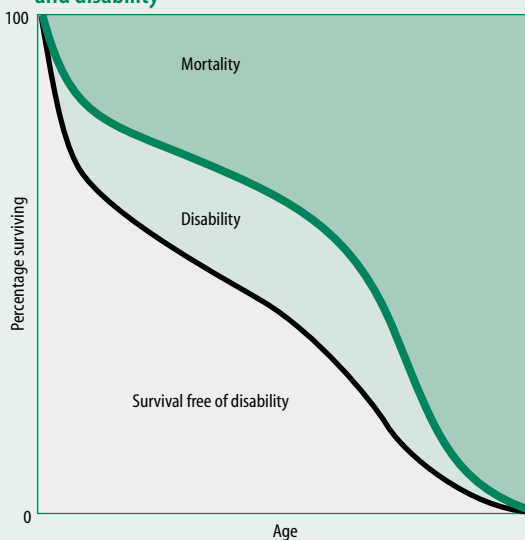
vival plus part of that for disability.

DALE is estimated from three kinds of information: the fraction of the population surviving to each age, calculated from birth and death rates; the prevalence of each type of disability at each age; and the weight assigned to each type of disability, which may or may not vary with age. Survival at each age is adjusted downward by the sum of all the disability

effects, each of which is the product of a weight and the complement of a prevalence (the share of the population not suffering that disability). These adjusted survival shares are then divided by the initial population, before any mortality occurred, to give the average number of equivalent healthy life years that a newborn member of the population could expect to live.

One important difference between the burden of disease estimation using disability-adjusted life years (DALYs) and that of DALE is that the former do, but the latter do not, distinguish the contribution of each disease to the overall result. DALE has the advantage that it does not require as many choices of parameters for the calculation, and it is directly comparable to the more familiar notion of life expectancy without adjustment.

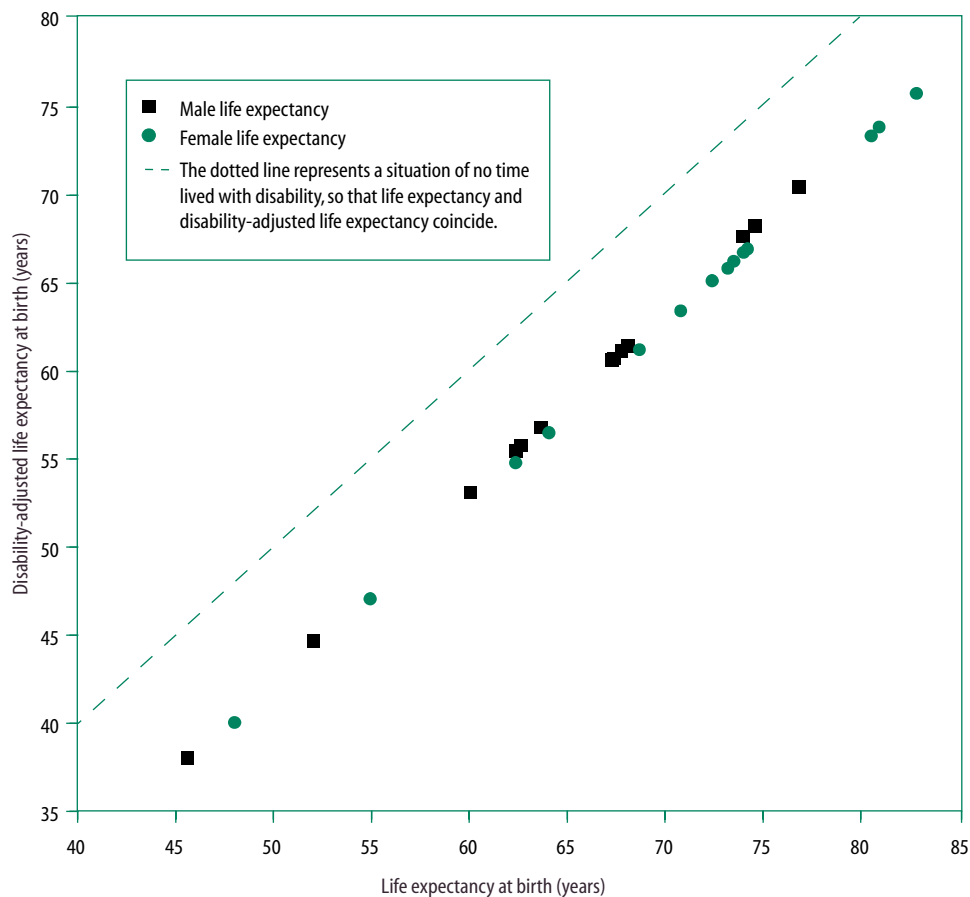
Summarizing population health from mortality and disability



Source: Murray CJL, Salomon JA, Mathers C. *A critical examination of summary measures of population health*. Geneva, World Health Organization, 1999 (GPE Discussion paper No. 12).

having no expectation of surviving infancy. *If everyone had the same life expectancy, adjusted for disability, the system would be perfectly fair with respect to health, even though people would actually die at different ages.* For a small number of countries it has been possible to estimate the distribution of life expectancy within the population using information on both child

**Figure 2.2 Life expectancy and disability-adjusted life expectancy for males and females, by WHO Region and stratum defined by child mortality and adult mortality, 1999**

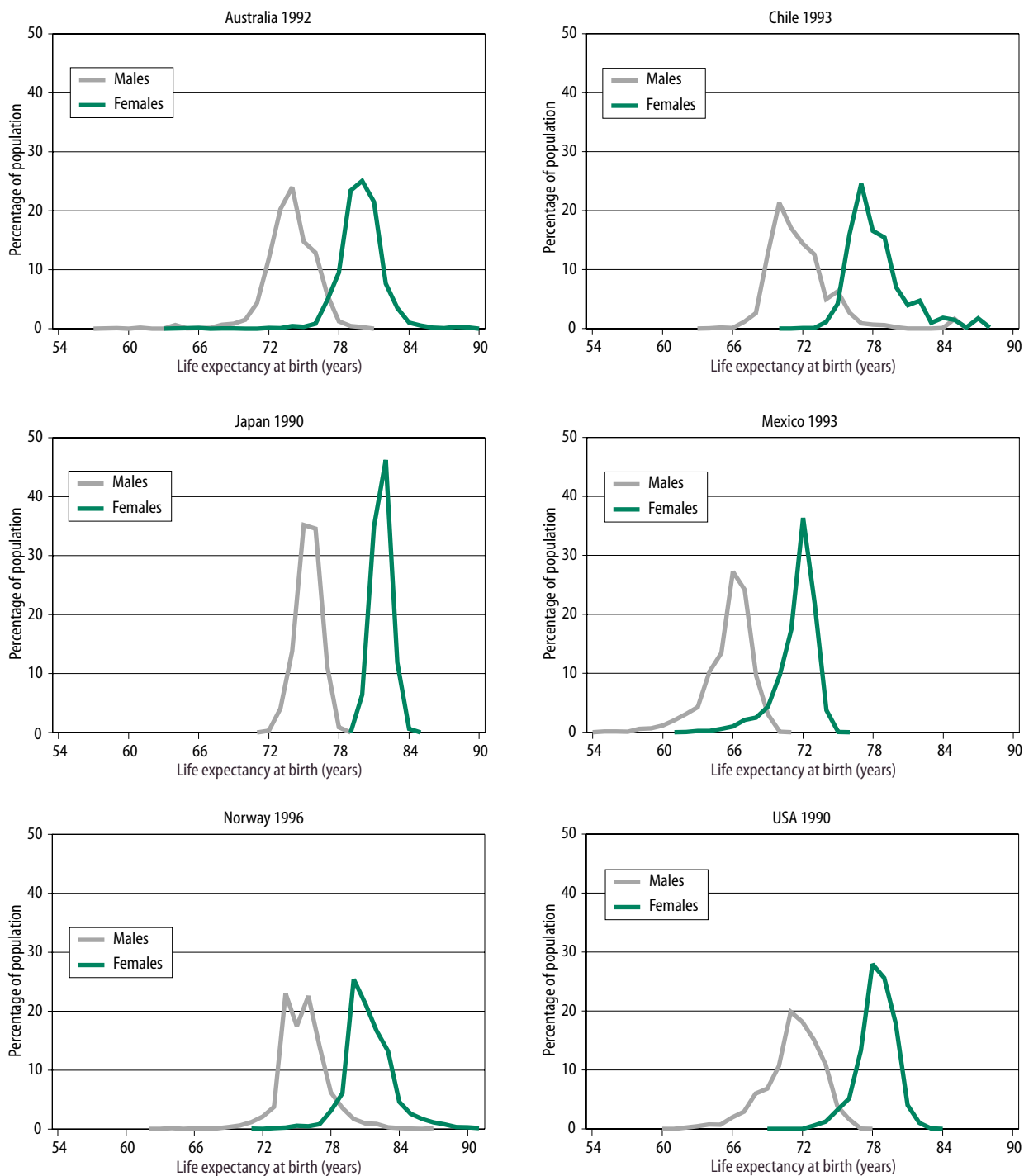


WHO Region	Mortality stratum		Males		Females	
	Child	Adult	Life expectancy	Disability adjusted	Life expectancy	Disability adjusted
AFR	High	High	52.0	44.6	54.9	47.0
	High	Very high	45.6	38.0	48.0	40.0
AMR	Very low	Very low	73.9	67.5	80.4	73.2
	Low	Low	67.3	60.6	74.1	66.8
	High	High	63.6	56.7	68.6	61.1
EMR	Low	Low	67.7	61.0	70.7	63.3
	High	High	60.0	53.0	62.3	54.7
EUR	Very low	Very low	74.5	68.1	80.8	73.7
	Low	Low	67.3	60.6	73.9	66.6
	Low	High	62.3	55.4	73.4	66.1
SEAR	Low	Low	67.2	60.5	73.1	65.7
	High	High	62.6	55.7	64.0	56.4
WPR	Very low	Very low	76.7	70.3	82.7	75.6
	Low	Low	68.0	61.3	72.3	65.0



and adult mortality; these results are presented below. For most countries, however, it has so far been possible to use only child mortality data. Because high-income countries have largely eliminated child mortality, the highest ranking countries in Annex Table 5 nearly all have relatively high incomes; most are European. A few Latin American countries which

Figure 2.3 Inequality in life expectancy at birth, by sex, in six countries



have made great progress in controlling child deaths also show considerable equality of health. Except for Afghanistan and Pakistan, all the countries ranked lowest on child health equality are in sub-Saharan Africa, where child mortality is still relatively high. When more complete data are available on inequalities in adult mortality they will be used in future WHO estimates, and these rankings will change, because high income countries differ more in adult than in child death rates.

Inequalities in life expectancy persist, and are strongly associated with socioeconomic class, even in countries with quite good health status on average (5). Figure 2.3 illustrates these inequalities for six countries, showing the distribution of life expectancy at birth for both men and women, using data on adult as well as child mortality, estimated from large numbers of small-area studies which cover the entire country. Among these six countries, health is most equally distributed in Japan. Both distributions of life expectancy are sharply peaked, concentrating the whole population of either men or women in a range of only about six years. There is far more inequality in Mexico and in the United States, and in both cases that arises because part of the population has a much lower expectation than the rest, after age five years. The inequality is particularly marked for men. An opposite pattern characterizes Chile, which shows very high equality in child health: the degree of adult inequality is about the same as for Mexico and the United States, but it arises because part of the population has an unusually high life expectancy. Australia and Norway both show more symmetric distributions. These results emphasize the value of judging health system achievement not only by averages or overall levels but by seeing whether everyone has about the same expectation of life.

*Responsiveness* is not a measure of how the system responds to health needs, which shows up in health outcomes, but of how the system performs relative to non-health aspects, meeting or not meeting a population's expectations of how it should be treated by providers of prevention, care or non-personal services. (The last category is least important, since individuals normally do not come into personal contact with such interventions. However, even public health measures such as vector control can be conducted with more or less respect for people and their wishes. Assessing the responsiveness of providers of non-personal services is a particular challenge.)

Some systems are highly unresponsive. The Soviet health system prior to 1990 had become highly impersonal and inhuman in the way it processed people. A common complaint in many countries about public sector health workers focuses on their rudeness and arrogance in relations with patients (6, 7). Waiting times for non-emergency surgery vary considerably among industrialized countries (8) and are the subject of much criticism of ministries of health (9). Recognizing responsiveness as an intrinsic goal of health systems establishes that these systems are there to serve people, and involves more than an assessment of people's satisfaction with the purely medical care they receive.

The general notion of responsiveness can be decomposed in many ways. One basic distinction is between elements related to respect for human beings as persons – which are largely subjective and judged primarily by the patient – and more objective elements related to how a system meets certain commonly expressed concerns of patients and their families as clients of health systems, some of which can be directly observed at health facilities. Subdividing these two categories leads to seven distinct elements or aspects of responsiveness.

Respect for persons includes:

- Respect for the dignity of the person. At the extreme, this means not sterilizing individuals with a genetic disorder or locking up people with communicable diseases, which would violate basic human rights. More generally, it means not humiliating or demeaning patients.
- Confidentiality, or the right to determine who has access to one's personal health information.
- Autonomy to participate in choices about one's own health. This includes helping choose what treatment to receive or not to receive.

Client orientation includes:

- Prompt attention: immediate attention in emergencies, and reasonable waiting times for non-emergencies.
- Amenities of adequate quality, such as cleanliness, space, and hospital food.
- Access to social support networks – family and friends – for people receiving care.
- Choice of provider, or freedom to select which individual or organization delivers one's care.

In general, responsiveness contributes to health by promoting utilization, but that is not always the case. Greater autonomy can mean that people do not take up an intervention because they perceive the individual benefit to be small or the risk to be substantial, and do not value the collective or population benefit. This is particularly likely for immunization, especially if there is fear of adverse reactions. Individual freedom to choose whether or not to be immunized is in conflict with the public health objective of high coverage to prevent epidemics. Such conflict has occurred, for example, in the United Kingdom for pertussis and in Greece for rubella vaccine (10). The overall performance of a health system may therefore involve trade-offs among objectives.

Opinions on how well a health system performs on such subjective dimensions as responsiveness might be influenced by any of a number of features of the systems themselves, or of the respondents. Since poor people may expect less than rich people, and be more satisfied with unresponsive services, measures of responsiveness should correct for

### Box 2.2 How important are the different elements of responsiveness ?

The key informant survey, consisting of 1791 interviews in 35 countries, yielded scores (from 0 to 10) on each element of responsiveness, as well as overall scores. A second, Internet-based survey of 1006 participants (half from within WHO) generated opinions about the relative importance of the elements, which were used to combine the element scores into an overall score instead of just taking the mean or using the key informants' overall responses.

Respondents were asked to rank the seven elements in order of importance, and the weights were derived from the frequencies with which an element was ranked first, second, and so on. Respect for persons and client orientation were rated as equally important overall, and the three elements of respect for persons were also regarded as all about equally important. The four elements of

client orientation received different rankings and therefore unequal weights. The final weights are shown in the table.

Analysis of the element scores themselves, as estimated by the key informants, showed three consistent biases: for the same country, women respondents gave lower scores than men, and government officials gave higher scores than more independent informants; and all informants' scores tended to be higher for countries with less political freedom, as measured by a composite index. The data were adjusted to make the scores comparable across countries by removing the influence of these factors, so that all the scores are estimates of the ratings that would be given in a politically free country, by respondents who did not work for the government, half of whom were women.

<i>Respect for persons</i>	
Total	50%
Respect for dignity	16.7%
Confidentiality	16.7%
Autonomy	16.7%
<i>Client orientation</i>	
Total	50%
Prompt attention	20%
Quality of amenities	15%
Access to social support networks	10%
Choice of provider	5%

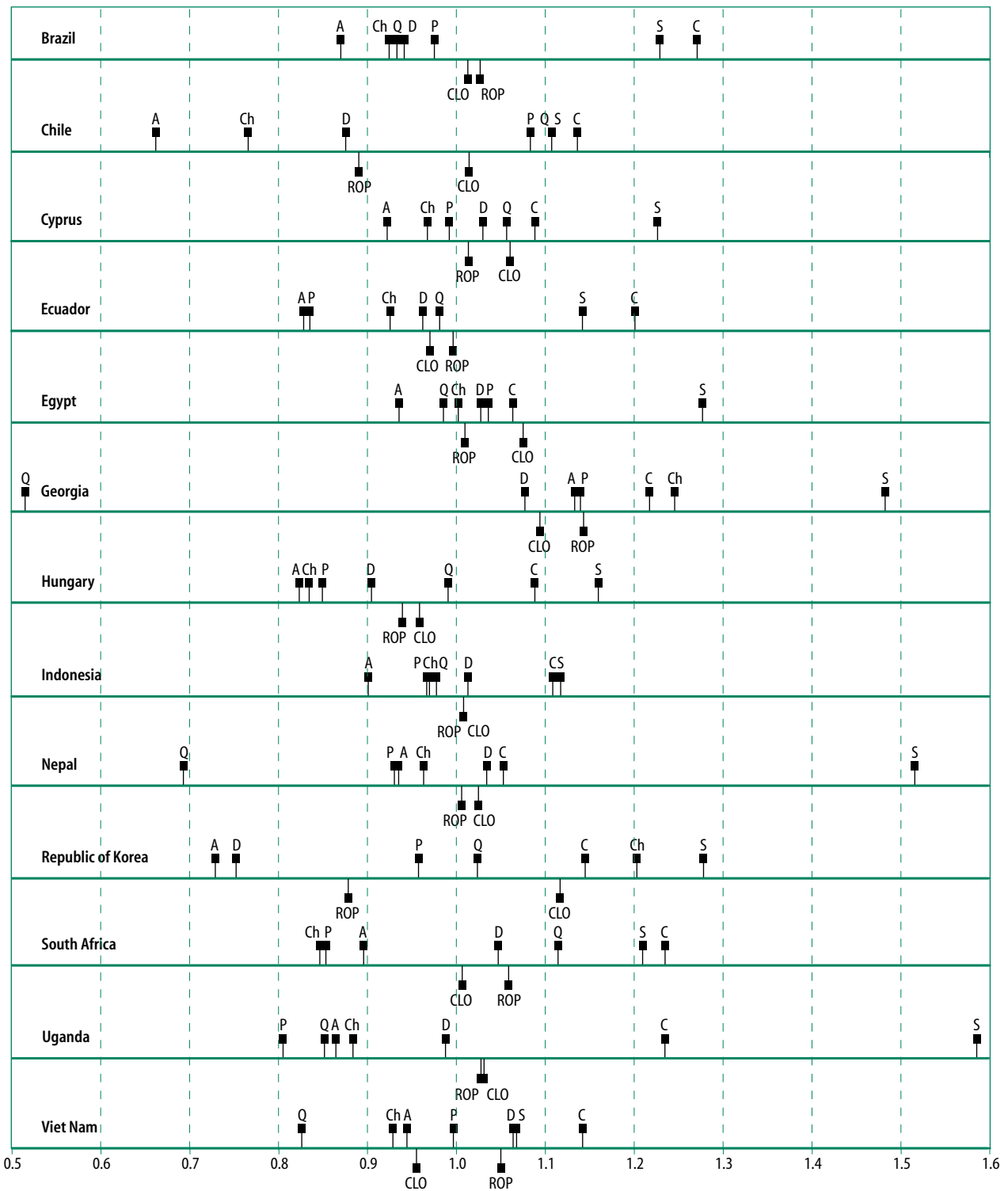
these differences, as well as for cultural differences among countries (11). Even without such adjustment, comparisons of how knowledgeable observers rate health system achievements can reveal on which aspects of responsiveness a system seems to satisfy its users best. Judgements about average level and inequality of the components of responsiveness were developed in each of 35 countries by a network of 50 or more key informants. A separate survey of over a thousand respondents was used to develop weights for combining these scores into an overall rating. Box 2.2 describes the results of this exercise. Estimates for other Member States were derived from the 35 observations, adjusted for differences among countries and informant groups. Surveys of population opinion and direct observation of health provision can both be used to complement these judgements.

Figure 2.4 illustrates in detail the scores of the seven individual elements, relative to the overall score, within each of 13 countries chosen to reflect all WHO Regions and typical of the entire set of countries studied. The health systems examined always appear to perform relatively well on the two dimensions of access to social support networks and confidentiality, sometimes very much better than on other aspects. The systematically high rating for social support may reflect a trade-off against the quality of amenities, because a health care facility that cannot, for lack of resources, offer good quality food or non-medical attention can compensate for that by allowing relatives and friends to attend to patients' needs. One reason why confidentiality seems not to be a problem in these countries may be that there is little private insurance and therefore little risk of coverage being denied because a provider reveals some information about a patient. There is somewhat less consistency at the other end of the scale, but autonomy is among the three lowest-rated elements of responsiveness 34 times out of 35 – and the lowest ranked element almost half the time – and performance is also often poor with respect to choice of provider and promptness of care.

As with health status, it is not only overall responsiveness that matters, if some people are treated with courtesy while others are humiliated or disdained. *A perfectly fair health system would make no such distinctions, and would receive the same rating of responsiveness on every element, for every group in the population.* In almost every country where key informants were surveyed, the poor were identified as the main disadvantaged group. In particular, they were considered to be treated with less respect for their dignity, to have less choice of providers and to be offered poorer quality amenities than the non-poor. In nearly as many cases, rural populations – among whom the poor are concentrated – were regarded as being treated worse than urban dwellers, suffering especially from less prompt attention, less choice of providers and lower quality of amenities. Some respondents in one or several countries also identified women, children or adolescents, indigenous or tribal groups or others as receiving worse treatment than the rest of the population.

The elements of client orientation, where the poor and the rural population are less well treated, all have economic implications: it generally costs more to assure quick attention and to offer high quality food, more space and well-kept facilities. It also makes cost control harder if people are allowed to choose their providers, and costs differ among them. The strongest associations occur for quality of basic amenities and promptness of attention. The former is closely related to income per head and to the share of private expenditure in total health spending; the latter is closely related to average years of schooling of the population, which is also associated with income. In contrast, the elements of respect for persons can be costless, apart perhaps from some training of providers and administrators. These elements – respect for dignity, autonomy, and confidentiality – show no relation to health system spending. There is scope for improving health system performance in these respects without taking any resources away from the primary objective of better health. This

Figure 2.4 Relative scores of health system responsiveness elements, in 13 countries, 1999



All scores are normalized relative to the average overall country responsiveness score = 1.

A = autonomy, C = confidentiality, Ch = choice of provider or facility, D = dignity, P = promptness, Q = quality of basic amenities, S = access to social support networks, ROP = respect of persons, CLO = client orientation.

is particularly the case for autonomy, where performance is often poorly rated.

Annex Table 6 reports adjusted scores for overall responsiveness, as well as a measure of fairness based on the informants' views as to which groups are most often discriminated against in a country's population and on how large those groups are. Either a larger group being affected, or more informants agreeing on that group's being treated worse than some others, implies more inequality of responsiveness and therefore less achievement of fairness. Since some elements of responsiveness are costly, it is not surprising that most of the highest ranked countries spend relatively large amounts on health. They are also often countries where a large share of provision is private, even if much of the financing for it is public or publicly mandated. However, the association with a country's income or health expenditure is less marked than it is for health status. Several poor African and Asian countries rank fairly high on the level of responsiveness. And countries that perform well on average for responding to people's expectations may nonetheless rank much lower on the distributional index.

*Fair financing* in health systems means that the risks each household faces due to the costs of the health system are distributed according to ability to pay rather than to the risk of illness: a fairly financed system ensures financial protection for everyone. A health system in which individuals or households are sometimes forced into poverty through their purchase of needed care, or forced to do without it because of the cost, is unfair. This situation characterizes most poorer countries and some middle and high income ones, in which at least part of the population is inadequately protected from financial risks (12).

Paying for health care can be unfair in two different ways. It can expose families to large *unexpected* expenses, that is, costs that could not be foreseen and have to be paid out of pocket at the moment of utilization of services rather than being covered by some kind of prepayment. Or it can impose *regressive* payments, in which those least able to contribute pay proportionately more than the better-off. The first problem is solved by minimizing the share of out-of-pocket financing of the system, so as to rely as fully as possible on more predictable prepayment that is unrelated to illness or utilization. The second is solved by assuring that each form of prepayment – through taxes of all kinds, social insurance, or voluntary insurance – is progressive or at least neutral with respect to income, being related to capacity to pay rather than to health risk.

Out-of-pocket payments are generally regressive but they can, in principle, be neutral or progressive. When this happens, and out-of-pocket expenses are not too large, they need not impoverish anyone or deter the poor from obtaining care. However, of all the forms of financing they are the most difficult to make progressive. Arrangements that exempt the destitute from user fees at public facilities, or impose a sliding scale based on socioeconomic characteristics, are attempts to reduce the risk associated with out-of-pocket payments (13, 14). Except when private practitioners know their clientele well enough to discriminate among them in fees – and the better-off accept that their charges will subsidize the worse-off – such arrangements are limited to public facilities, which often account for only a small share of utilization in poor countries. And even then, such schemes require relatively high administrative costs to distinguish among users, and typically affect only a small amount of total risk-related payments.

For this reason, financial fairness is best served by more, as well as by more progressive, prepayment in place of out-of-pocket expenditure. And the latter should be small not only in the aggregate, but relative to households' ability to pay. Prepayment that is closely related to *ex ante* risk, as judged from observable characteristics – risk-related insurance premiums,

for example – is still preferable to out-of-pocket payment because it is more predictable, and may be justified to the extent that the risks are under a person's control. However, the ideal is largely to disconnect a household's financial contribution to the health system from its health risks, and separate it almost entirely from the use of needed services. The question of how far insurance prepayments may be related to risks, and how such premiums should be financed, including subsidies for those unable to pay, is treated in Chapter 5.

*Ex post*, the burden of health financing on a particular household is the share that its actual health expenses are of its capacity to pay. The numerator includes all costs attributable to the household, including those it is not even aware of paying, such as the share of sales or value-added taxes it pays on consumption, which governments then devote to health, and the contribution via insurance provided, and partly financed, by employers.

The denominator is a measure of the household's capacity to pay. In poor households, a large share goes for basic necessities, particularly food, whereas richer households have more margin for other spending, including spending on health care. Food spending is treated as an approximation to expenditure on basic needs. Total non-food spending is taken as an approximation of the household's discretionary and relatively permanent income, which is less volatile than recorded income (15) and a better measure of what a household can afford to spend on health and other non-food needs.

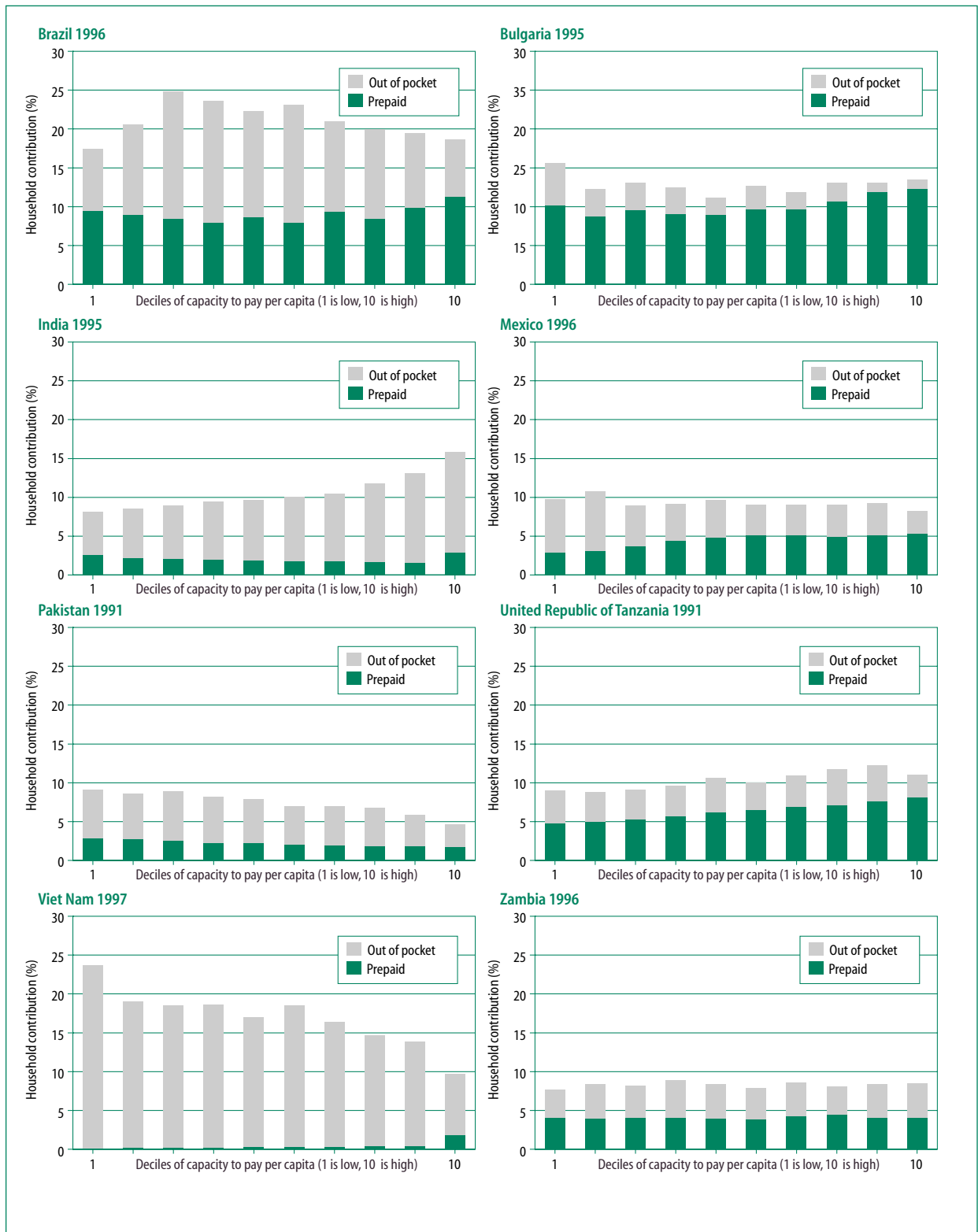
In sum, *the way health care is financed is perfectly fair if the ratio of total health contribution to total non-food spending is identical for all households, independently of their income, their health status or their use of the health system.* This indicator expresses the trenchant view of Aneurin Bevan, that "The essence of a satisfactory health service is that the rich and the poor are treated alike, that poverty is not a disability, and wealth is not advantaged." (16). Clearly the financing would be unfair if poor households spent a larger share than rich ones, either because they were less protected by prepayment systems and so had to pay relatively more out of pocket, or because the prepayment arrangements were regressive. But to identify fairness with equality means that the system is also regarded as unfair if *rich* households pay more, as a share of their capacity. Simply by paying the same fraction as poor households, they would be subsidizing those with lower capacity to pay. It is true that well-off households might choose to pay still more, particularly by buying more insurance, but that can be considered equitable only if the extra spending is prepaid and if the choice is entirely voluntary and not determined by the system of taxes or mandatory insurance contributions.

Families that spend 50% or more of their non-food expenditure on health are likely to be impoverished as a result. Detailed household surveys show that in Brazil, Bulgaria, Jamaica, Kyrgyzstan, Mexico, Nepal, Nicaragua, Paraguay, Peru, the Russian Federation, Viet Nam and Zambia more than 1% of all households had to spend on health half or more of their full monthly capacity to pay, which means that in large countries millions of families are at risk of impoverishment. Invariably the reason is high out-of-pocket spending. This high potential for financial catastrophe has much to do with how the health system is financed, and not only with the overall level of spending or the income of the country.

The fairness of the distribution of financial contribution is summarized in an index which is inversely related to the inequality in the distribution, and presented in Annex Table 7. The index runs from zero (extreme inequality) to 1 (perfect equality). For most countries, and particularly for most high income countries, the value is not far from 1, but great inequality characterizes a few countries in which nearly all health spending is out-of-pocket, notably China, Nepal and Viet Nam. However, in some countries where most spending is out-of-



**Figure 2.5 Household contributions to financing health, as percentage of capacity to pay, in eight countries**



pocket, there is nonetheless little inequality because that spending is relatively progressive and few families spend as much as half their non-food expenditure on health. Bangladesh and India are examples. Generally, high values of equality are associated with predominantly prepaid financing, but Brazil shows extreme inequality despite a high share of prepayment, because of the great inequality in incomes and the large number of families at risk of impoverishment.

The summary measure of fairness does not distinguish poor from rich households. Figure 2.5 introduces this distinction, by showing how the burden is distributed across deciles of capacity to pay, and divided between prepayment and out-of-pocket spending, in eight low and middle income countries. Prepayment is clearly progressive – the rich contribute a larger share – in Mexico and the United Republic of Tanzania, and also in Bangladesh and Colombia (not shown). It is actually regressive in India and Pakistan, and also in Guyana, Kyrgyzstan, Nepal, Peru and the Russian Federation (not shown). In other countries – Brazil, Bulgaria, Jamaica, Nicaragua, Paraguay, Romania and Zambia – the prepaid contribution is distributed more or less neutrally or varies irregularly. Out-of-pocket spending shows more variation, as might be expected; for example, it is progressive in India and quite regressive in Pakistan and Viet Nam, where there is almost no prepaid financing at all.

Total non-food spending also includes whatever the household spends out of pocket on health care. That spending is largely unpredictable or transitory, so to include it may overstate the family's capacity to pay. If out-of-pocket expenditure is small, it makes no difference; but if it is large, it may have been financed by selling assets, going into debt, requiring more family members to work or for some to take on more hours or other employments, or even temporarily reducing consumption of necessities. If household capacity to pay is defined as non-food spending *less* out-of-pocket health spending, then families with large out-of-pocket expenditures are classified as poor, instead of being scattered throughout the population. The way the health system is financed then looks systematically less fair in most countries, and the culprit is always the large share of out-of-pocket spending. Pre-

### Box 2.3 What does fair contribution measure and not measure ?

The way fair contribution to health care finance is measured is strictly *ex post*, referring to what households actually contribute rather than to their *ex ante* risks of needing health care. That means that there is no need to estimate the "coverage" of the population by different risk-sharing schemes. Coverage in financial rather than nominal terms – how much people are really protected, not simply whether they have insurance or participate in social security – is hard to estimate beforehand, and in any case such coverage is, like accessibility, an instrumental rather than a final goal. Nominal coverage does provide people with a sense of security which also

affects their spending and saving decisions, but that is not a goal in itself.

People who do not use care when they need it, because they cannot afford the out-of-pocket cost, appear to spend less than they really need to. Estimating what they *would* have spent if they could afford it would give a different distribution of contributions, and would almost surely show even more unfairness. Basing the measure on what is actually spent – which is all that the data allow – overstates the degree to which a health system achieves a fair distribution of the financing burden.

The measure also says nothing about how a family obtains the cash

to pay out of pocket for health care (or for some forms of prepayment such as "health cards" or vouchers). Households much of whose income is in kind rather than cash may forego health care because they cannot obtain the cash when needed, and the data will show only that they did not spend. Without further analysis there is no way to distinguish illiquidity from all the other reasons why a health need did not eventuate in expenditure. A less serious but sometimes still consequential liquidity problem arises when a household has to pay out of pocket for care, and then wait for reimbursement from an insurer. This need to finance care temporarily arises for populations wealthy

enough to have formal – usually private – insurance.

Finally, and most important, fair financing means only equity in how the financial burden of supporting a health system is shared. It says nothing about whether the *utilization* of health services is fair, which is an equally crucial issue in the overall fairness of the system. Fair financing is concerned with the principle of *from each according to ability*, but not with the principle of *to each according to need*. Unfairness in use relative to need shows up in inequalities in health status, because service utilization ought to reduce such inequalities so far as they are amenable to intervention.

payment in low income countries is commonly too small a share of the total to offset the regressive and very unequal impact of out-of-pocket spending. What is worse, in many countries there is *no* offsetting effect because prepayment via taxes is also regressive. In those cases the poor thus suffer twice – all of them have to pay an unfair share whether or not they use health services, and then some of them have also to pay an even more unfair contribution out of pocket. These are the strongest findings to emerge from the analysis of financing; their implications are developed further in Chapter 5. Box 2.3 discusses four other features of this way of measuring fairness in financing.

Much of the analytical effort behind this report went into developing more and better information about expenditure on health and constructing national health accounts. As described further in Chapter 6, knowledge of where resources are coming from, through what channels they flow and how they are used, is crucial to better stewardship of the system. Annex Table 8 presents the estimates of total health spending, its separation into private and public sources, the distinction between tax-financed and social security health spending in the public sector, and that between insurance and out-of-pocket spending in the private sector, and the overall distinction between prepayment and out-of-pocket spending that helps determine how fairly health systems are paid for. These data, besides being of direct interest, have been used to check the estimates of household expenditure discussed above and to estimate values for indicators that are strongly related to spending.

Whatever the sources and distribution of finance, the level of resources devoted to health is an input into the system, not an outcome: it is what makes the outcomes possible, and against which the system's achievements should be evaluated. The next two sections take up the question of how best to do this, first by developing an overall measure of attainment and then by relating that achievement to resource use, as a measure of performance.

#### Box 2.4 Weighting the achievements that go into overall attainment

To derive a set of weights for the different achievements that compose overall attainment, WHO conducted a survey of 1006 respondents from 125 countries, half from among its own staff. The questions were designed to elicit not only views about how important each goal is relative to the others (for example, responsiveness compared to health status), but also opinions about what kind of inequality matters most. The responses were checked for consistency and bias, and yielded nearly identical values in each of many different groups – poorer versus richer countries, men versus women, WHO staff versus other respondents. The final weights are shown in the table.

As expected, health is regarded as the most important of the objectives, clearly the primary or defining goal of a system. But fully half of the concern for health is a concern for equality, not simply for a high average. Taking “health” apart into two goals emphasizes the great value of fairness, and not only of goodness. This is fully consistent with WHO's concentration on the poor, the least healthy, the worst-off in society. Equal weights also result from the survey for the overall level and for distribution or equality where responsiveness is concerned. In total, how the system treats people in non-health aspects is as important as either health level or health

equality. And fairness in how health is paid for, which is not a major traditional concern of WHO or the ministries of health it deals with and supports, receives the relatively large weight of one-fourth, equal to that for responsiveness. Both in this case and in that of responsiveness, the weight assigned by respondents probably reflects the direct or intrinsic importance of the objective, and also the indirect or instrumental contribution it makes to achieving good health; it is difficult to separate these two aspects. There is clear agreement that a well-functioning health system should do much more than simply promote the best possible level of overall health.

The exercise of weighting the five objectives also provides values for the relative importance of goodness and fairness. Together, the *levels* of health and of responsiveness receive a weight of three-eighths of the total. The three distributional measures, which together describe the *equity* of the system, account for the remaining five-eighths. Countries which have achieved only rather short life expectancies and cannot adequately meet their peoples' expectations for prompt attention or amenities may nonetheless be regarded as having health systems which perform well with respect to fairness on one or more dimensions.

<i>Health (disability-adjusted life expectancy)</i>	
Total	50%
Overall or average	25%
Distribution or equality	25%
<i>Responsiveness</i>	
Total	25%
Overall or average	12.5%
Distribution or equality	12.5%
<i>Fair financial contribution</i>	
Distribution or equality	25%

## OVERALL ATTAINMENT: GOODNESS AND FAIRNESS COMBINED

To the extent that a health system achieves a long disability-adjusted life expectancy, or a high level of responsiveness (or a high degree of equality in either or both), or a fair distribution of the financing burden, it can be said to perform well with respect to that objective. Since a system can do well on one or more dimensions and poorly on others, comparison across countries or through time requires that the five goals be summed into a single overall measure. There is no natural scale on which to add together years of life, responsiveness scores, and measures of inequality or fairness, so combining the measures of achievement means assigning a weight or relative importance to each one. Box 2.4 describes the procedure and the results.

Applying these weights to the achievements described in Annex Tables 5, 6 and 7 yields an overall attainment score for each health system. These scores are presented in Annex Table 9, together with an estimate of the uncertainty around each value, derived from the uncertainties for the components. Because rich countries generally enjoy good health, and because high incomes allow for large health expenditures which are also predominantly prepaid and often largely public, the ranking by overall attainment is closely related to income and health spending. However, the large weight given to distributional goals explains why, for example, Japan outranks the United States and why Chile, Colombia and Cuba outrank all other Latin American countries. It is not surprising that, with three Asian exceptions, the 30 worst-off countries are all in Africa.

## PERFORMANCE: GETTING RESULTS FROM RESOURCES

The overall indicator of attainment, like the five specific achievements which compose it, is an absolute measure. It says how well a country has done in reaching the different goals, but it says nothing about how that outcome compares to what might have been achieved with the resources available in the country. It is *achievement relative to resources* that is the critical measure of a health system's performance.

Thus if Sweden enjoys better health than Uganda – life expectancy is almost exactly twice as long – that is in large part because it spends exactly 35 times as much per capita on its health system. But Pakistan spends almost precisely the same amount per person as Uganda, out of an income per person that is close to Uganda's, and yet it has a life expectancy almost 25 years higher. This is the crucial comparison: why are health outcomes in Pakistan so much better, for the same expenditure? And it is health expenditure that matters, not the country's total income, because one society may choose to spend less of a given income on health than another. Each health system should be judged according to the resources actually at its disposal, not according to other resources which in principle could have been devoted to health but were used for something else.

Health outcomes have often been assessed in relation to inputs such as the number of doctors or hospital beds per unit of population. This approach indicates what these inputs *actually* produce, but it tells little about the health system's *potential* – what it could do if it used the same level of financial resources to produce and deploy different numbers and combinations of professionals, buildings, equipment and consumables. In these comparisons, the right measure of resources is money, since that is used to buy all the real inputs.

To assess relative performance requires a scale, one end of which establishes an upper

limit or “frontier”, corresponding to *the most that could be expected of a health system*. This frontier – derived using information from many countries but with a specific value for each country – represents the level of attainment which a health system might achieve, but which no country surpasses. At the other extreme, a lower boundary needs to be defined for *the least that could be demanded of the health system* (17). With this scale it is possible to see how much of this potential has been realized. In other words, comparing actual attainment with potential shows how far from its own frontier of maximal performance is each country’s health system.

WHO has estimated two relations between outcomes and health system resources. One estimate relates resources only to average health status (disability-adjusted life expectancy, DALE), which makes it somewhat comparable to many previous analyses of performance in health. The other relates resources to the overall attainment measure based on all five objectives. The same value of total resources is used for a country in both cases, because there is no way to identify expenditure as being directed to producing health services, determining responsiveness or making the financing more or less fair. The same is true of resources used to improve the distribution of health or responsiveness, rather than the average level.

Each frontier is a function of one other variable besides health system expenditure. That is the average years of schooling in the adult population, which is a measure of human

### Box 2.5 Estimating the best to be expected and the least to be demanded

WHO’s estimates of the upper and lower bounds of health system performance differ in two important ways from most analyses of what health systems actually achieve. The first is that a “frontier” is meaningful only if no country can lie beyond it, although at least one must lie on it. The frontier or upper limit is therefore estimated by a statistical technique which allows for errors in one direction only, minimizing the distances between the frontier and the calculated performance values. (The lower bound is estimated by the conventional technique of allowing errors in either direction.) The second is that the object is not to *explain* what each country or health system has attained, so much as to form an estimate of what should be possible. The degree of explanation could be increased by introducing many more variables. If tropical countries show systematically lower achievement in health, because of the effects of many diseases concentrated near the equator, a variable indicating

tropical location would raise the explanatory or predictive power. Similarly, if outcomes are worse with respect to equality in ethnically diverse countries, a variable reflecting that heterogeneity would explain the outcomes observed.

The difficulty with the attempt to explain as much as possible is that it leads to a different frontier, according to every additional variable. There would be one for tropical countries and another for colder climates; one for ethnically mixed countries and another for those with more uniform populations; and so on. If performance were measured relative to the frontier for each type of country, almost every health system might look about equally efficient in the use of resources, because less would be expected of some than of others. Every additional explanation would be the equivalent of a reason for not doing better. This is particularly true of explanations related to individual diseases: AIDS and malaria are major causes of health loss in many sub-Saharan African countries, but to include

their effects in the estimation of the frontier means judging those countries only according to how well they control all *other* diseases, as though nothing could be done about AIDS and malaria. This is the reason for estimating the frontier according to nothing but expenditure and human capital, which is a general measure of society’s capacity for many kinds of performance, including performance of the health system.

The measures of attainment draw on data referring to the past several years, to make the estimates more robust and less susceptible to anomalous values in any one year. The measures of expenditure and human capital are similarly constructed from more than one year’s data. Nonetheless, both the outcomes and the factors that determine potential performance are meant to describe the current situation of countries. They do not take into account how past decisions and use of resources may have limited what a system can actually achieve today – which could also be a rea-

son for poor performance – nor do they say how quickly a poorly performing system might be expected to improve and come closer to the frontier.

This way of estimating what is feasible bypasses two particularly complex issues which are well illustrated by control of tobacco-related mortality and disability. One is that many actions taken by health systems produce results only after a number of years, so that resources used today are not closely related to outcomes today. If a health system somehow persuaded all smokers to quit and no one to take up the habit, it would be many years before there was no more tobacco-induced disease burden.<sup>1</sup> The other is that no health system could reasonably be expected to bring smoking prevalence down to zero any time soon, no matter how hard it tried. Determining how to evaluate progress rather than only a health system’s current performance is one of many challenges for future effort.

<sup>1</sup> Jha P, Chaloupka F, eds. *Tobacco control policies in developing countries*. Oxford, Oxford University Press for the World Bank and the World Health Organization, 2000.

capital and therefore of the long-run potential, if not the current or actual, state of development of the country. It is a proxy for most of the factors outside the health system that contribute to health status, and probably also to the degree of responsiveness and to how health is financed. Box 2.5 explains how the upper and lower limits are estimated and how they should be interpreted.

Since the estimation is based entirely on country data rather than a model of what is ideal or feasible, and since there are upper limits to all the achievements, the frontier rises rapidly with additional resources when spending is low, and then rises more and more slowly as expenditure reaches the levels typical of rich countries. A health system can move *towards* the frontier by improving performance, that is by achieving more with the same resources. It can move *along* the frontier by spending more or less on health and reaching a different level of attainment but the same degree of performance. The entire frontier can also move *outward*, as new knowledge makes it possible to achieve better health or other outcomes, for given health system resources and a given level of human capital. Most of the enormous improvement in health over the last century and a half, described in Chapter 1, is due to such an expansion or outward movement of what it is possible to achieve.

If there were no health system in the modern sense, people would still be born, live and die; life expectancy would be much less than now, but it would not be zero. There would be no expenditure on health and hence no question of how fairly the financial burden was distributed. Similarly, there would be no responsiveness. So the minimum level of achievement would involve only health status, and in the absence of information about inequalities, only the average level of health. In the measure of overall attainment the values for the other four objectives, including all those related to inequality, would be set at zero. To estimate this minimum, WHO has used information from a limited number of countries *circa* 1900, relating life expectancy – with no adjustment for disability – to estimates of income. The situation at the turn of the last century is taken as the starting point for the great advances made possible by increased knowledge, investment and resources devoted to health. Some of the changes have the effect of raising the minimum – the eradication of smallpox is the best example. The emergence of HIV/AIDS and of tobacco-related disease have the opposite effect, making it harder than it was in 1900 to achieve a given level of health.

The question for any health system today is, given the country's human capital and the resources devoted to its health system, how close has it come to the most that could be asked of it? Relating outcomes in this way to the estimated minimum and maximum attainments and to the use of economic resources defines the overall indicator of system *performance*: to perform well means to move away from the minimum attainment and come close to the maximum. In economic terms, performance is a measure of efficiency: an efficient health system achieves much, relative to the resources at its disposal. In contrast, an inefficient system is wasteful of resources, even if it achieves high levels of health, responsiveness and fairness. That is, it could be expected to do still better, because countries spending less do comparably well or countries spending a little more achieve much better outcomes.

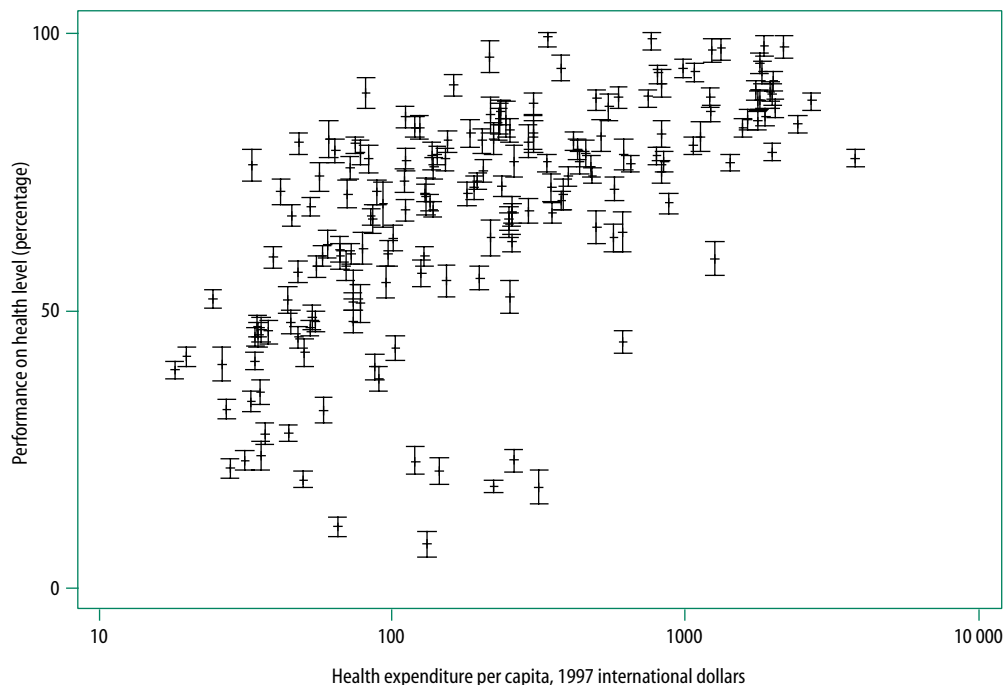
Annex Table 10 presents two indicators of health system performance. The first is based only on the average health status in disability-adjusted life expectancy (DALE) presented in Annex Table 5, comparing the frontier for that objective alone to a country's resource use and human capital. In this case, the upper and lower bounds between which performance lies are strictly comparable, and the measure can be compared to other estimates of what determines health outcomes. As with the measures of attainment, these values carry estimates of uncertainty. Figure 2.6 shows the estimated distribution of performance for all

countries with respect to DALE. Higher health expenditure is associated with better health outcomes, even when performance is judged relative to expenditure rather than absolutely. Very poor countries evidently suffer from other handicaps than low spending and low educational attainment. The few countries where spending is below \$10 per person per year seldom appear to achieve more than 75% of the life expectancy that should be possible, whereas most countries spending more than \$1000 achieve at least 75% of the possible. Higher spending is also associated with less variation in performance. Disturbingly large variations in life expectancy relative to spending and education occur at low and middle levels of expenditure where there is the greatest need to understand and reduce differences in achievement. A large part of the explanation is the HIV/AIDS epidemic: the 25 worst-off countries are all African nations suffering from a severe burden of AIDS. (Box 2.5 explains why the epidemic was not taken into account in defining the frontier of the possible.)

The second indicator in Annex Table 10 is based on the overall attainment measure presented in Annex Table 9 and assesses performance relative to the frontier defined for all five elements of achievement. The intervals around these values are much larger than for DALE alone because of the uncertainty surrounding the other components. These components also account for some considerable changes in the ranking, but the best performing systems still seem to be those of relatively rich countries and the worst off are predominantly poor and in Africa. Figure 2.7 presents the distribution of overall performance, which shows somewhat less variation than Figure 2.6: countries that perform poorly with respect to health alone sometimes compensate for this by doing better in responsiveness or financing or in dealing with health inequality. Nonetheless the rankings of the two performance measures are rather closely associated, with a small number of countries that do much better by one measure than by the other.

The belief that the system should be accountable for the level and distribution of attainment on the goals of health, responsiveness and fair financing, all relative to health

**Figure 2.6 Performance on level of health (disability-adjusted life expectancy) relative to health expenditure per capita, 191 Member States, 1999**





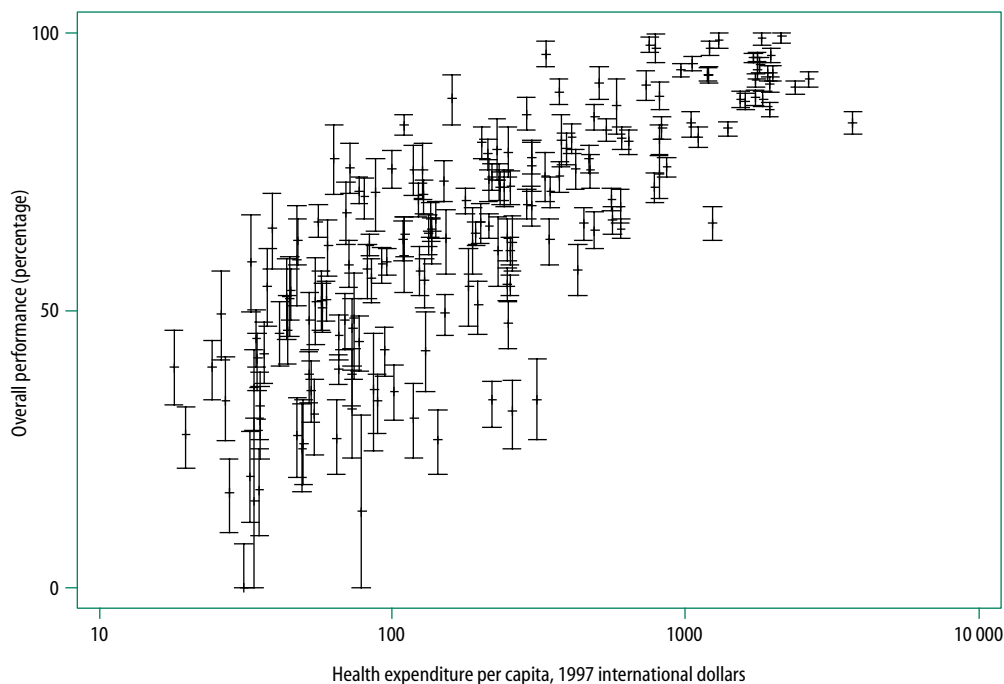
expenditure, will remain central in WHO's work to support health systems development over the coming years. From this issue, each year's *World health report* will contain more complete and better measures of countries' achievements, and WHO will support countries to strengthen local skills to analyse and improve health system attainment and performance.

## IMPROVING PERFORMANCE: FOUR KEY FUNCTIONS

Policy to improve performance requires information on the principal factors which explain it. Knowledge of the determinants of health system *performance*, as distinct from understanding of what determines health *status*, remains very limited. This report focuses on a few universal functions which health systems perform, as indicated in Figure 2.1 above, asking what it means for those functions to be discharged well or poorly and suggesting how they are associated with differences in achievement among countries. This helps to look at the health system overall, rather than building up from the component sub-systems, organizations or programmes, as is more common in evaluations of performance (18).

The service provision function is the most familiar, and in fact the entire health system is often identified with just service delivery. The classification here emphasizes that providing services is something the system *does*; it is not what the system *is*. Much of what is included in the financing function occurs outside what is usually considered to be the health system, as a process which happens to collect revenues and put them at the system's disposal. Treating fairness in financial contribution as one of the intrinsic goals of the system requires viewing the function partly as another of the tasks that the system *does*, rather than pas-

**Figure 2.7 Overall health system performance (all attainments) relative to health expenditure per capita, 191 Member States, 1997**



sively receiving money from somewhere else. It is the system which collects some of the funds directly, pools all that are pooled except for general taxation, and purchases goods and services. This means the system is at least partly accountable to society for how resources are raised and combined, and not only for how they are ultimately used.

Every health system makes some investments in creating resources, but these also are sometimes regarded as coming from outside the system itself. In the short run, the system can only use the resources created in the past, and often can do little to change even how they are employed. But in the long run, investment also is something the system does – and precisely because investments are long-lived, it has a responsibility to invest wisely. Relating achievements to total system expenditure may show that a system is performing badly precisely because what can be obtained from today's resources is needlessly limited by how resources were invested yesterday and the day before.

The fourth function is called *stewardship*, because the concept is well described by the dictionary definition: *the careful and responsible management of something entrusted to one's care* (19). People entrust both their bodies and their money to the health system, which has a responsibility to protect the former and use the latter wisely and well. The government is particularly called on to play the role of a steward, because it spends revenues that people are required to pay through taxes and social insurance, and because it makes many of the rules that are followed in private and voluntary transactions. It also owns facilities on trust from the citizens. Private insurers and practitioners, however, perform this function in only a slightly restricted degree, and part of the state's task as the overall steward or trustee of the system is to see to it that private organizations and actors also act carefully and responsibly. A large part of stewardship consists of regulation, whether undertaken by the government or by private bodies which regulate their members, often under general rules determined by government. But the concept embraces more than just regulation, and when properly conducted has a pervasive influence on all the workings of the system.

These functions are identifiable in widely differing health system structures (20, 1). At one extreme is a system in which functions are substantially combined in a single organization which raises, pools and allocates funds to a fairly monolithic group of service providers who are its own employees. The Norwegian health system resembles this type of structure, as did the British National Health Service prior to 1990. A system may instead have a high degree of "vertical" segmentation. Separate organizations such as the ministry of health, social security funds, the armed forces, charitable organizations, or private insurers may pay their own providers, raise and allocate funds and provide services, for non-overlapping populations. The health systems of much of Latin America bear some resemblance to this model, although patients often get care from two or more of the vertically separate organizations. A system could also have "horizontal" integration of each function – one organization performing it – but a different organization for each function. No system quite corresponds to this, because there is never a single bloc of providers, unless they are part of a fully integrated system. However, some systems such as that of Chile separate collection and pooling for a large share of resources, and employ a large number of providers under a single organization. At the opposite extreme from a monolithic organization is a system with separate institutions raising funds and paying providers under pluralistic provision arrangements in which few providers "belong" to the financing institution. The Colombian system, following the reforms introduced since 1993, looks somewhat like the latter.

Chapters 3, 4, 5 and 6 concentrate in turn on key characteristics of each of the four functions – service delivery, investment, financing and stewardship – and on some factors

affecting performance, examining patterns in countries at different income levels. The financing function obviously is most important for the goal of fairness in paying for the system, but how it is carried out also affects health outcomes and even has some effect on responsiveness. The service delivery function is most tied to health outcomes, but also matters greatly for responsiveness. And stewardship affects everything.

## REFERENCES

1. Murray CJL, Frenk J. *A WHO framework for health system performance assessment*. Geneva, World Health Organization, 1999 (GPE Discussion Paper No. 6).
2. Roemer MI. *National health systems of the world*. New York, Oxford University Press, 1991.
3. Centers for Disease Prevention and Control, Centro Latino-Americano de Investigación en Sistemas de Salud, and Pan American Health Organization. *Public health in the Americas*. Washington, DC, Pan American Health Organization, 1999 (unpublished document).
4. *World development report 1993 – Investing in health*. Washington, DC, The World Bank, 1993.
5. Acheson D et al. *Independent inquiry into inequalities in health*. London, The Stationery Office, 1998.
6. Gilson L, Alilio M, Heggenhougen K. Community satisfaction with primary health care services: an evaluation undertaken in the Morogoro region of Tanzania. *Social Science and Medicine*, 1994, **39**(6): 767–780.
7. Bassett MT, Bijlmakers L, Sanders DM. Professionalism, patient satisfaction and quality of health care: experience during Zimbabwe's structural adjustment programme. *Social Science and Medicine*, 1997, **45**(12): 1845–1852.
8. Hurst J. Challenges to health systems in OECD countries. *Bulletin of the World Health Organization*, 2000, **78**(6) (in press).
9. Donelan K et al. The cost of health system change: public discontent in five nations. *Health Affairs*, 1999, **18**(3): 206–216.
10. King S. Vaccination policies: individual rights v. community health. *British Medical Journal*, 1999, **319**: 1448–1449.
11. de Silva A. *A framework for measuring responsiveness*. Geneva, World Health Organization, Global Programme on Evidence for Health Policy, 1999 (unpublished paper).
12. Fabricant S, Kamara C, Mills A. Why the poor pay more: household curative expenditures in rural Sierra Leone. *International Journal of Health Planning and Management*, 1999, **14**: 179–199.
13. Nolan B, Turbat V. *Cost recovery in public health services in sub-Saharan Africa*. Washington, DC, Economic Development Institute of The World Bank, 1995.
14. Bennett S, Creese A, Monasch R. Health insurance schemes for people outside formal sector employment. Geneva, World Health Organization, 1998 (Current Concerns, ARA paper No. 16, document WHO/ARA/CC/98.1).
15. Friedman M. *Theory of the consumption function*. Princeton, NJ, Princeton University Press, 1957.
16. Bevan A. *In place of fear*. London, Heinemann, 1952.
17. Donabedian A, Wheeler JRC, Wyszewianski L. Quality, cost, and health: an integrative model. *Medical Care*, 1982, **20**: 975–992.
18. Cumper GE. *The evaluation of national health systems*. Oxford, Oxford University Press, 1991.
19. *Meriam Webster's deluxe dictionary*. Pleasantville, New York/Montreal, Reader's Digest, 1998.
20. Murray CJL, Kreuser J, Whang W. Cost-effectiveness analysis and policy choices: investing in health systems. *Bulletin of the World Health Organization*, 1994, **72**(4): 663–674.