HEALTH SYSTEMS PERFORMANCE ASSESSMENT

Christopher J.L. Murray, DPhil, MD

Annual Lecture 2003

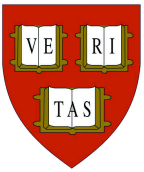
Research OHE Consulting
Office of Health Economics

www.ohe.org
ISBN 1 899040 23 4
September 2005

HEALTH SYSTEMS PERFORMANCE ASSESSMENT

Professor Christopher J.L. Murray,
DPhil, MD

OHE Annual Lecture 2003



The Office of Health Economics
12 Whitehall London SW1A 2DY
Telephone: +44 (0)20 7930 9203
www.ohe.org

About the Author

Christopher Murray is Richard Saltonstall Professor of Population Policy, the Director of the Harvard University Initiative for Global Health and the former Executive Director of the Evidence and Information for Policy Cluster at the World Health Organization. A physician and health economist, his early work focused on tuberculosis control and the development of the pioneering Global Burden of Disease project at Harvard University. Recently, he has initiated major new approaches to the measurement of population health, cost-effectiveness analysis and the conceptualization, measurement and national application of health systems performance assessment. He has authored or edited seven books, many book chapters and more than 90 journal articles in internationally peer-reviewed publications.

Office of Health Economics

The Office of Health Economics (OHE) was founded in 1962. Its terms of reference are to:

- commission and undertake research on the economics of health and health care;
- collect and analyse health and health care data from the UK and other countries;
- disseminate the results of this work and stimulate discussion of them and their policy implications.

The OHE is supported by an annual grant from the Association of the British Pharmaceutical Industry and by revenue from sales of its publications, consultancy and commissioned research.

Contents

Introduction

I would like to thank Adrian Towse, the Office of Health Economics and Professor Culyer for the opportunity to present this lecture. I am very honoured to be asked to speak in the UK about the work on health systems performance assessment (HSPA) at the World Health Organization. In the next hour I would like to give you some reflections on this work and trace some implications for the UK. These reflections are based on five years of work involving a large number of researchers and policy analysts at WHO and in academic institutions around the world.

In this presentation I would like to cover six topics:

1. The political economy of WHO's work on health systems performance assessment.
2. The conceptual framework for health systems performance assessment including its evolution after *The World Health Report 2000*.
3. Systematic attempts to fill key information gaps with new data collection efforts.
4. Some of the remaining methodological challenges and some small steps in the direction of new methods that may help in the task ahead.
5. New empirical findings obtained over the last few years.
6. Final reflections.

HEALTH SYSTEMS PERFORMANCE ASSESSMENT

The Political Economy of Performance Assessment

Why did Dr Brundtland, the Director General of WHO launch a programme of work at the Organization on health systems and on health systems performance when she took office five years ago? Her goals in 1998 are the goals that we have today. In the short run, we would like to help WHO's Member States undertake three critical tasks. First, health systems performance assessment (HSPA) should enable decision-makers use evidence to inform planning and strategic decision-making such as capital investments or new programme development. Second, HSPA should help managers monitor health systems so that they can respond in real-time with needed corrections. Third, HSPA should allow decision-makers and society to monitor the progress of the health system towards agreed upon targets and goals.

These three short-term objectives for HSPA are also intimately related to a longer-term goal, namely, building up the evidence base on what works and what does not for health systems. I think everybody here, economists, policy analysts, managers, and clinicians, are aware of the influence that the evidence-based medicine movement has had. The widespread use of randomized clinical trials and other types of studies to inform clinical decision-making has fundamentally changed the discourse on clinical medicine. Unfortunately, the evidence base to understand what attributes of health system design or organization lead to better performance is much weaker.

We believe that the health systems evidence base will only be strengthened in the long run if there are some common tools, frameworks, and measurement approaches. These common approaches can serve the short-term objectives for HSPA and can also support the long-term development of an evidence base on what works and what does not for health systems. Building the evidence base

on health systems, however, will require patience. It is a 10, 15 or 20 year agenda and not a five year one.

Three years ago, in June 2000, we published *The World Health Report (WHR): Health Systems: Improving Performance.* That was neither the beginning nor the end, but a first instalment of our work on health systems. The WHR entailed two components. The first was a substantive discussion on what was known or believed at the time about four different aspects of health systems: their financing; how services should be organized; how human, physical, and intellectual resources should be generated; and the critical dimension of stewardship of health systems.

Second, the WHR also included rankings for 191 countries for four different outcomes that health systems contribute to:

1. population health measured using healthy life expectancy and inequalities in child mortality;
2. responsiveness of health systems, the systems analogue to interpersonal quality of care or patient experience;
3. fairness in financial contribution; and, perhaps most ambitiously,
4. efficiency of using inputs to improve these outcomes.

As we all know, there has been a wide array of reactions to *The World Health Report 2000*. Many responses were very detailed and insightful scientific commentaries published in the literature or presented at scientific conferences and meetings convened by WHO. I will return to these scientific issues later in more detail. In addition to this scientific discourse, a number of other types of reactions have influenced the way the discussion about HSPA has taken place.

Two debates in particular, more strategic than scientific have arisen. People's position on these topics is often valuable. The first debate is on the use of composite indicators and the presentation of composites in league tables. Given the UK discussions on hospital league tables, this is familiar terrain in this country.

Second, there has also been a debate on the role of WHO. Many see the Organization as the private counsellor to the Ministry of Health, while others see it as a global monitor of outcomes. The former leads WHO to foster close ties to the current Minister of Health and his/her leadership team. Such an intimate and confidential relationship between WHO and the Minister enhances the Organization's ability to communicate important technical and political messages to governments. A different perspective is that WHO should act as the health conscience for the world, putting information out to the public, and the media, and trying to influence or change health policy debates through the provision of information. This debate on counsellor versus monitor continues to this day; maintaining both roles may be extremely difficult for the Organisation.

Two other aspects of reactions to *The World Health Report 2000* must be noted. There was a strong philosophical or ideological response, not so much to the rankings of health systems, but to the actual substantive content of the report, particularly of Chapter 3 on health services. Some commentators believed that there was a coded message encouraging private sector provision in the report. Therefore, the supporters of universal public sector provision of services felt that this was undermining their views.This concern has been particularly strong in Latin America.

There has also been a wide array of "political opportunism," both on the positive and the negative side. A number of ministers and heads of state saw the results of these rankings as an opportunity to drive forward agendas for change. In selected other cases, individuals considered the rankings as challenging their own political careers, which gave rise to a complicated set of reactions.

How has WHO responded to the full range of these concerns? Dr Brundtland announced at the WHO Executive Board in January 2001 a very in-depth review process of the work on health systems performance assessment. This included regional meetings where Member States in each of WHO's six regions were invited along with technical experts to discuss performance assessment and how it should be improved. The review also included eight academic conferences on specific topics. In addition, WHO has undertaken a major programme of work to expand the empirical basis for performance

assessment, as well as to tackle some of the methodological issues that had been raised in the responses to *The World Health Report 2000*.

These multiple channels of consultation and work were overseen by two important groups: a small advisory group responsible for the process of review, chaired by Professor Mahmoud Fathalla, who was the Chairman of WHO's most senior advisory group, the Advisory Committee on Health Research; and a scientific peer review group, chaired by Professor Sudhir Anand from Oxford University. This elaborate, detailed, and rich process of consultation is well-documented in a book published by WHO, *Health Systems Performance Assessment: Debates, Methods and Empiricism*. This book includes in its entirety the report of the Scientific Peer Review Group.

Conceptual Framework

Based on these deliberations, the WHO conceptual framework for health systems performance assessment has evolved and I think improved. It comprises four broad categories for thinking about health systems:

1. Inputs;
2. Functions of health systems. These were outlined in *The World Health Report* and have not changed much in terms of how to think about systems' organization. The four functions are financing; the generation of human, physical, and intellectual resources; the provision of services; and the notion of the guiding vision and regulatory functions for the government, namely stewardship.
3. Health systems should contribute to key final outcomes. In this framework final outcomes are the average level of population health, inequalities in health, responsiveness - our term for patient experience - and the distribution of payment to health care systems, which we call "fairness in financial contribution."
4. Finally, it is desirable that systems use inputs to contribute to outcomes in an efficient manner.

What is particularly new is a strong focus that emerged through our regional and technical consultations: performance assessment should take advantage of, and use information on the delivery of effective interventions to people who need them - coverage in the sense of immunization coverage, or coverage of cervical cancer screening. 'One aspect of having good metrics of effective coverage is to be able to solve two of the major problems of efficiency estimation: the issue of casual attribution - how do we attribute final outcomes, particularly to the action of the health system - and how do we deal with the obvious time lags between certain interventions (smoking cessation for example) and health outcomes which may be 20, 30 or 40 years later?

Effective coverage as the mediating factor between personal and non-personal health services and health outcomes is itself influenced by many factors, both from a demand and supply side, or a household and provider perspective. Part of the work in the conceptual framework is to try and understand what the contribution of different factors is: e.g. the price of drugs, the price of health care, the gap between perceived and actual need, physical distance, cultural acceptability, resource availability, technical quality of providers, adherence, etc.

In fact, coverage also permeates into the way we think efficiency should be measured. One aspect of having good metrics of effective coverage is to be able to solve two of the major problems of efficiency estimation: the issue of causal attribution - how do we attribute final outcomes, particularly to the action of the health system, and how do we deal with the obvious time lags between certain interventions - smoking cessation, for example - and health outcomes, which may be 20, 30 or 40 years later? If we have good information on effective coverage we believe that at least these two aspects of efficiency estimation can be addressed. There is still the question of how to figure out that relationship (or "efficiency frontier") and I will come back to it later.

Another aspect of the development of this framework has been the attempt to think in a coherent way of extending a system-level framework down to the level of the provider, so at least the metrics and the approaches are conceptually coherent. We have made an effort,

particularly for hospitals but also for other providers, to think about their ability to deliver interventions. In fact we have defined, at least for measurement purposes, the quality of providers as the fraction of potential health gain delivered to a particular set of patients. This issue will be further discussed later.

Expanding the Empirical Base

Our conceptual framework has clearly expanded. I believe it is richer and more likely to yield insights into policies to improve performance if we are able to actually operationalize it. Part of the challenge is better data. The paucity of good data was one of the overwhelming criticisms of *The World Health Report 2000*. Some have argued that there was an overuse of imputation methods and the available data for other countries were too weak. Taking these criticisms to heart, we have invested in a major effort to improve the empirical base for performance assessment.

There are two thrusts to expand the empirical base: better use of existing data and the development and application of standardized data collection instruments across multiple countries. Existing data sets can be used much more effectively for performance assessment. For example, consider household expenditures on health. Most countries have income and expenditure surveys. It is very rare that ministries of health have either the access to the data, which are often held by national statistical offices, or the capacity to analyse the data. To address this problem, we have worked with ministries of health in 104 countries to gain access to micro data and help them analyse details of health expenditure at the household level for about 148 different household surveys. Likewise, we have tried to develop insights into human resource inputs, not only at the national, but also at the district level using census micro data. That work is underway in 17 countries. For the measurement of child health inequality, we have undertaken a much more extensive use of existing DHS and PAPCHILD surveys.

The main thrust of our work on expanding the empirical basis for performance assessment has been the development of new survey instruments and the fielding of household surveys in three phases:

First, in a learning phase, we have developed new instruments for measuring responsiveness, health status, poverty, and effective coverage, and we have explored the trade-off between data quality and cost in different survey modes including face-to-face, telephone, and self-administered surveys. We undertook, in the years 2000 and 2001, 71 surveys in 61 countries, comprising of a mixture of face-to-face interview surveys, self-administered postal surveys, and computer-assisted telephone interviews. There are more surveys than countries because we conducted dual surveys in some countries to test mode effects. As it turns out, 28 out of 30 OECD countries were included in what was called the MCSS (Multi-Country Survey Study). I will be showing you some empirical results for the UK using the survey undertaken in 2001 as part of this survey programme.

The World Health Survey was a first step in the direction of developing what I consider a much improved platform for data collection, particularly for performance assessment, and also for monitoring the Millennium Development goals. It has been intended to create modules or build on existing modules on poverty measurement, health insurance, health expenditure, health status, health state valuation, responsiveness, coverage of interventions, including pharmaceutical use, mortality, cause of death, health occupations and social capital. The instrument was tested in 12 countries in 2002 - the 12 Country Pilot Study. The World Health Survey final instrument is now in the field in 73 countries in many languages. We hope that the data will be made public with a minimal time lag. Putting them in the pubic domain will be an essential step to encourage national researchers as well as governments to use this type of information.

The World Health Survey is distributed across all regions with a good mixture of poor countries (18 sub-Saharan African countries), a strong participation of European Union countries, as well as a number of other high-income countries, such as Australia. Despite many attempts to convince the national authorities, the World Health Survey has not been undertaken in the US. When such comparative data on the US become available, a very interesting set of comparisons across OECD countries will be facilitated.

Methods Development

Part of the challenge for HSPA is improving the empirical base, but many of the technical issues that emerged during the debates before and after *The World Health Report 2000* have been about methods. Many years of research will be required in order to address all of these issues. I believe that in eight specific areas there has been reasonable progress in the last three years. I will briefly mention these areas, giving more detailed illustrations for two of them.

1. For the estimation of mortality by age and sex in countries with limited information, and because of the HIV epidemic, we have developed a new system of model life tables based on the Brass Logit Life Table System, called the Modified Logit Life Table System. The new method reflects the fact that there are patterns of mortality now in parts of the world, such as Eastern Europe and countries with an HIV epidemic, not captured in previous model life tables.
2. The comparability of self-reported items for responsiveness, health, social capital, etc. has been improved through the anchoring vignette method and the associated statistical models. I will describe these methods in more detail later.
3. Statistical models for looking at total inequality of mortality risk - that is both the within-group and between group-variation in mortality risk - have also been advanced.
4. A critical component to measuring coverage of interventions is to be able through household surveys to identify those in need. So if you want to know what fraction of people with angina is receiving appropriate treatment, you have to first identify those who have angina. There have been some real advances in methods and instruments to undertake probabilistic diagnosis in household surveys, helping to obtain the numerator, i.e. those people on appropriate treatment.
5. The approaches for assessing catastrophic and impoverishing health payments through household

income and expenditure surveys have been completely revised conceptually, definitionally, and methodologically.

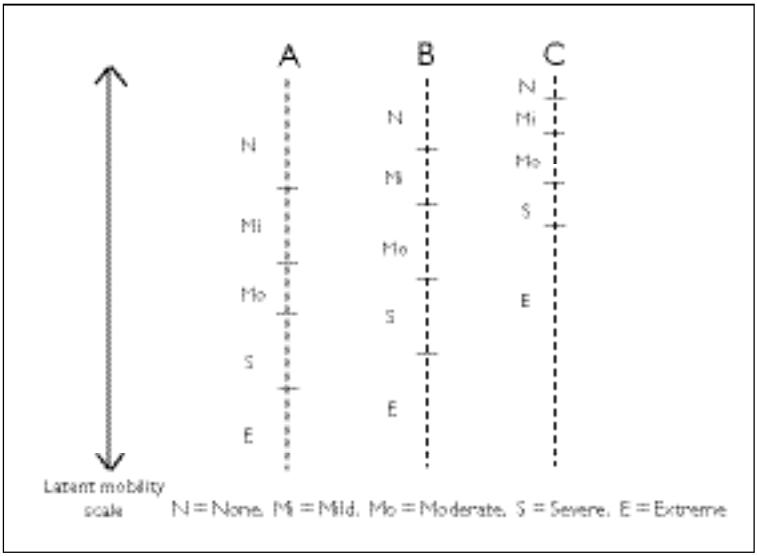
6. One of the many uses of our work on cost-effectiveness, which we call CHOICE (Choosing interventions that are Cost-Effective), is to directly assess the efficiency frontier. The methods for WHO-CHOICE include global guidelines for cost-effectiveness analysis, a standardized framework for looking at the population effects of health intervention (called PopMod), a standardized costing approach (CostIt), and an approach to stochastic league tables (McLeague).
7. Work on health state valuations using multiple methods is increasing our understanding of variation in the valuation function across individuals and cultures. We believe it to be a very important aspect for the future, although it is in an early phase of its development.
8. Finally, we are working on simple tools based on administrative data to look at hospital quality using new approaches to risk adjustment.

I will now give two illustrations regarding developments in items (2) and (8) above: anchoring vignettes and measuring hospital quality. Anchoring vignettes may have broad implications. Figure 1 shows the basic problem around self-reported items, whether it is for a domain of health or a domain of patient experience.

Consider the mobility for an individual. Mobility in this diagram is shown as an unobserved latent variable. We ask an individual about his/her mobility with a typical health instrument question: "Do you have difficulties moving around?" The response categories that the individual can use are: "no problems," "mild problems," "moderate problems," "severe problems" or "extreme problems." The five response categories mean that there are four transition points in terms of the latent variable mobility which represent the level of mobility that an individual will shift from using one response such as "no problems" to the next response category "mild problems." These

transition points are called cut-points and may be different for different individuals. In fact, there is absolutely no reason that all individuals will map from mobility into these response categories in the same way. Figure 1 shows three individuals or three groups of individuals (A, B or C) that map from latent mobility into the response categories using different cut-points.

Figure 1 **Response category cut-point shifts**



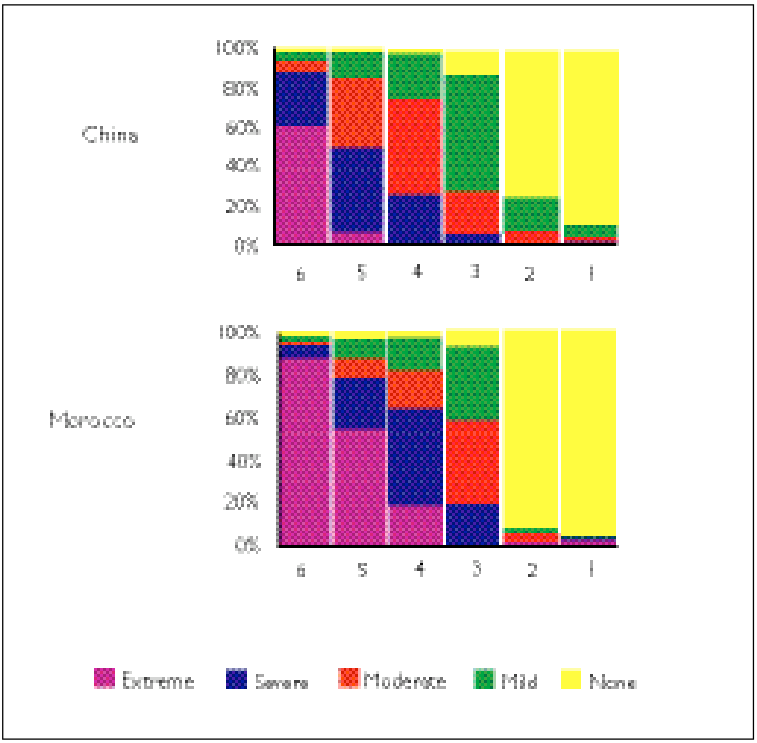
In psychometrics individual differences in cut-points are called Differential Item Functionin (DIF). We strongly believe based on our (DIF) empirical work that DIF is not only a theoretical issue, but also a common problem. There are systematic differences across individuals as well as across groups in how people use response categories. The differences are not because of translation problems, but because of more fundamental reasons. People's mappings from the latent variable one is trying to measure into a particular response category will differ. There is no reason why they would all be the same. The approach called Anchoring Vignettes tries to address this problem. It works by giving stories or vignettes to a subsample of respondents. For example, and referring back to the previous example illustrated in Figure 1, we present individuals with six mobility

vignettes. Individuals respond for each of these stories to the same question: "Does person A have difficulties moving around?" At one extreme, does Paul, who is a marathon runner, have difficulties moving around? At the other extreme, does David, the quadriplegic, have difficulties moving around? In between these two scenarios lie intermediate situations. From people's responses we can understand how they are using their response categories.

Figure 2 shows an illustration from survey data in China and Morocco, from the World Health Survey 12 Country Pilot Study.

Figure 2 shows the results from six vignettes. Number 6 in the horizontal axis is the quadriplegic, while number 1 is the marathon runner. The figure shows the distribution of individuals in each

Figure 2 **Illustrative example for China and Morocco**



country responding for that vignette using the five response categories ("no problems," "mild," "moderate," "severe" or "extreme"). Purple is "extreme," yellow is "no problems."

As expected, there is a gradient in both countries from the marathon runner to the quadriplegic. Respondents are more likely to say that quadriplegics have extreme problems and that the marathon runner has no problems. While the gradient is present, there is tremendous variation in the response category for any given vignette within a population. Individuals differ in how they rate the vignettes. Part of the differences is due to stochastic measurement error, but there is also a systematic element: people have different cut-points.

When we compare the distribution of responses in China and Morocco, it is clear that in Morocco people are more likely to use the categories "extreme" and "severe" for any given level of mobility, while in China people are more likely to use the categories "none" or "mild." This is evidence of systematic differences in cut-points between China and Morocco. In other words, differential item functioning means that we should not directly compare the responses to the item "Do you have problems moving around?" between China and Morocco.

There is an associated set of statistical models that have been developed to use and extract information from vignette data in order to correct individuals' responses so that they are on a common scale. This approach is called the Compound Hierarchal Ordered Probit Model and is more fully documented in the literature.

The second illustration of improved methods refers to cross-walking system-level notions of performance to provider levels of performance. WHO's work on this method is in its infancy, but we are hopeful that administrative data can be analysed to provide useful comparable insights on hospital quality in the future. We define the quality of a provider as the fraction of potential health gain that could be delivered to the set of patients seen by that provider. On the measurement side, we would like to test if administrative admission and discharge data with primary, secondary diagnoses and procedures coding can be used to approximate potential health gain that is being delivered. We have been working on risk adjustment methods and

microdata from three countries - USA, Australia, and Brazil - in a pilot study to see what is possible.

There is an extensive literature on quality measurement using both administrative and detailed clinical correlate data sets. It is therefore unlikely that WHO has got anything particularly new to say, except the Organization's ability to attempt to persuade a range of countries, particularly middle- and low-income countries, to use approaches to quantify provider quality.

The issues now discussed are based on the HCCUPS data from the US, which come from a sample of about 2,000 hospitals. The data include all of their admissions and discharge records from 1988-99. I will concentrate on data for respiratory and circulatory conditions, based on 800,000 and 1.7 million discharges respectively.

We plot data on hospital-specific relative risks for odds ratios after taking into account risk adjustment by age, income, place and primary and secondary diagnoses, and so on. Data show a wide range of statistically significant differences among hospitals, with relative risks as low as 0.5 for respiratory conditions, and relative risks as high as 4, or even higher in the outlier case. This information suggests that there is a four to eight-fold variation within the US in the probability of dying after a given risk-adjusted diagnosis. This is suggestive of a tremendous variation in quality, some of which is clearly relatable to hospital characteristics such as volume. However, the details of what might be the causes of this variation in quality are out of the scope of this lecture.

We believe that many quality issues are fundamentally related to hospital level attributes - the information system, staffing patterns, and management of the hospital. Of course there is a component linked to particular providers and particular services, but there is probably a lot of common variation or common influence of organizational aspects on provider quality.

An incidental finding which I have to comment on because I find it intriguing, although not centrally related to this presentation, is the observation that in the US there has been an apparent dramatic improvement in hospital quality over time. We can observe that for at

least circulatory and respiratory conditions, there is about a 40 per cent reduction in the risk-adjusted hospital death rate over the decade. This is not true for all causes. If we look at obstetric admissions there is almost no change, but for a number of major or complicated condition groups, this type of reduction is a fact. It is not due to sending people home just before they die, which might have been a plausible first reaction. The fraction of deaths in hospital does not seem to be changing over the period very much, or certainly not enough to explain the pattern we just described. Some of the improvement is technology-driven, but some may have been due to the emphasis in the US on quality metrics and performance improvement.

We believe that this approach to measure quality, while very simplistic, could be extremely strengthened by improving primary data collection at the level of the facility at a relatively low cost. If the construct here is to measure potential health gain, if we could actually add to the deaths in hospital information a short health status instrument at admission and a short health status instrument at discharge and, even better, at 30 or 90 days post-discharge, the power of the analysis would increase tremendously. This actually isolate the contribution in terms of health gain that a hospital is making, controlling for condition and other risk determinants. If this approach seems fruitful, WHO will certainly encourage this type of development of new information systems.

Some New Findings

In this part of the lecture, I will share with you some new empirical results that come from the methods development and their application to new data sets. I will follow the HSPA outcomes framework: levels of health, health inequalities, responsiveness, fairness in financial contribution, and then discuss overall health system efficiency.

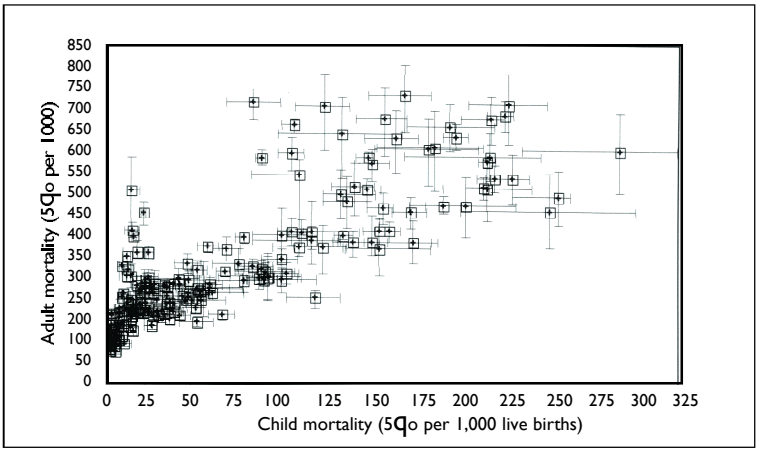
Health

At WHO we start with a very simple heuristic when we look at data concerning level of health: we plot child mortality on the x-axis and adult mortality on the y-axis. Child mortality is measured as the probability of death between birth and age 5, and adult male

mortality is measured as the probability that a 15 year old will die before his 60th birthday.

The figure illustrates a number of important points. First, in 2000 despite a century of progress in reducing child mortality, there is still an enormous range in child mortality across countries. Sierra Leone has child death rates of 279 per thousand, and Sweden of 4 per thousand. Second, for adult mortality the range is even larger. The probability of dying between ages 15 and 59 for males is as high as 703 per thousand in a country like Botswana and as low as 85 per thousand in Iceland. The countries with extremely high adult mortality all suffer a substantial HIV epidemic. There is also a large cluster of countries from the former Soviet Union and parts of Eastern Europe that have very high levels of adult male mortality for their level of child mortality.

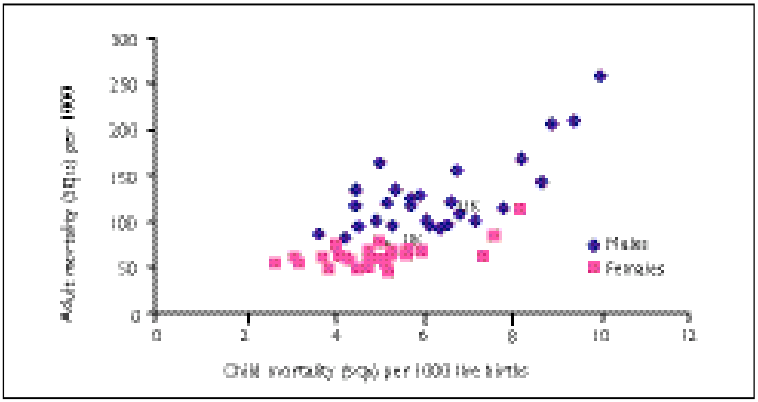
Figure 3 **Adult vs. child mortality, males, 191 Member States, 2000**



Even excluding these two groups of countries, there is great variation in levels of adult mortality for a given level of child mortality, from 150 to 270 per thousand, for example, and vice versa. The same level of adult male mortality is associated with a child mortality of 7 all the way up to 75. We have to recognize that there is heterogeneity in the fundamental epidemiological patterns in countries.

If we look at the same data for the OECD countries, we still observe a similar set of issues.

Figure 4 **2002 adult mortality vs. child mortality. OECD without Mexico and Turkey.**



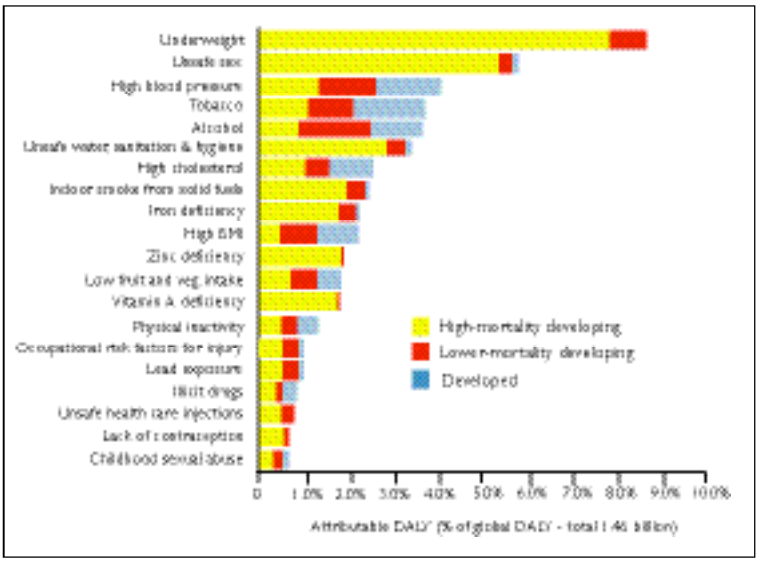
There is dramatically more heterogeneity across the OCD countries, at least in absolute terms, for adult male mortality than for child mortality. We suspect that there is much greater inequality within countries in adult male mortality than there is in child mortality or in adult female mortality. There is a stronger relationship between child mortality and adult female mortality. It is as if the key drivers for adult male health must include factors that play much less of a role for women and children.

Referring to the global scale, WHO annually provides a picture of the leading causes of the global burden of disease. In 2001, the number one cause of burden was conditions arising during the perinatal period, followed by lower respiratory infections as number two. Lower respiratory infections are a major cause of death for poor children in poor countries. The global pandemic of HIV has pushed this disease to be the third largest contributor to the global burden of disease. The fourth cause in 2001 is depression, a disabling condition rather than a condition causing mortality because suicide deaths are counted under violence. The remaining diseases and injuries in the list of the top 14 that make up 50% of the total burden of disease include diarrhoeal diseases, ischaemic heart disease and stroke, malaria, road traffic

accidents, tuberculosis, maternal conditions, chronic obstructive pulmonary disease, congenital anomalies and measles. At the global level, the 14 causes of burden that account for half all years of healthy life lost include a mixture of major infectious diseases, non-communicable diseases and injuries.

We have also extended this type of monitoring global epidemiology to risk factors. In last year's *World Health Report 2002* we had an assessment of the 20 leading risk factors worldwide. We measured the fraction of the global burden of disease that can be attributed to exposures to these risks. The leading risk factor is underweight, followed by unsafe sex, which account for respectively about 9 per cent and 6 per cent of the global burden. In Figure 5, the magnitude of each risk factor is represented in three colours. The harm in terms of the global burden of disease that occurs in high-mortality developing countries is shown in yellow, the harm in lower-mortality developing countries in red, and the harm in developed countries in blue.

Figure 5 **Attributable disease burden of 20 risk factors**



There are risks which mainly harm poor countries with high mortality. These include underweight, unsafe sex, unsafe water, sanitation and hygiene, indoor smoke from solid fuels, iron, zinc, and vitamin A deficiency. There are also a series of truly global risks. The distribution of the harm from these risks challenges the classic diseases of affluence paradigm. The burden from these risks is nearly equally distributed across high-mortality developing, lower-mortality developing, and developed countries. These global risks are high blood pressure, tobacco, alcohol, high cholesterol, obesity or high BMI, low fruit and vegetable intake, and physical inactivity.

The same analysis was done for the OECD leading risk factors and diseases. Although we do not have detailed data for the UK, I suspect that the pattern would be quite similar to the OECD average. The data OECD now discussed excludes Mexico, Turkey, Poland and Hungary, which belong fundamentally to different epidemiological zones. For burden from diseases and injuries, the number one cause in high-income countries is depression, followed by ischaemic heart disease and Alzheimer's, stroke and alcohol use. Hearing loss, which is new in this analysis because there are new data and many more prevalent surveys on it giving rise to a new appreciation of its magnitude as a cause of burden. Finally chronic obstructive pulmonary disease, lung cancers, osteoarthritis, and road traffic accidents are also included in the list of OECD most important diseases or injuries. The top 10 diseases and injuries in the OECD countries are much more focused on conditions that contribute to poor functional health status as opposed to causing premature mortality.

Figure 6 **OECD leading risk factors, diseases or injuries in 2002, (excluding Mexico, Turkey, Poland and Hungary**

Risk factor	% DALYS	Disease or injury	% DALYs
Tobacco	12.0	Unipolar depressive disorders	8.8
Blood pressure	7.3	Ischaemic heart disease	7.0
Alcohol	6.8	Alzheimer and other dementias	4.4
Overweight	6.7	Cerebrovascular diseases	4.4
Cholesterol	5.6	Alcohol use disorders	4.4
Physical inactivity	2.7	Hearing loss, adult onset	3.4
Low fruit and Veg	2.5	COPD	3.2
Illicit drugs	2.3	Trachea/bronchus/lung cancers	3.0
Unsafe sex	0.8	Osteoarthiritis	2.7
Iron deficiency	0.8	Road traffic accidents	2.5

On the risk factor side, number one in OECD countries is tobacco, followed by blood pressure, alcohol and obesity. The magnitude of blood pressure and alcohol is about 7 per cent of burden, dropping slightly down on cholesterol to about 5.6 per cent. Physical inactivity, low fruit and vegetable intake, and illicit drugs are about half the magnitude of those previous three.

At the country level, our metric of health status is healthy life expectancy (HALE). HALE is one single summary measure of health status at the population level. Healthy life expectancy can best be thought of as the expectation of life in full health. In Japan on average for men and women combined this is 74.8, followed by Sweden and Switzerland. Italy, rather interestingly, is at number five. Spain and France are in the list at positions seven and ten respectively. The UK is 19th with a healthy life expectancy of 70.7. The gap in terms of healthy life expectancy is still quite large between the UK and Japan, suggesting that there is a lot of scope for the UK vis-à-vis a country like Sweden or Japan to make progress in improving non-fatal health outcomes and reducing mortality.

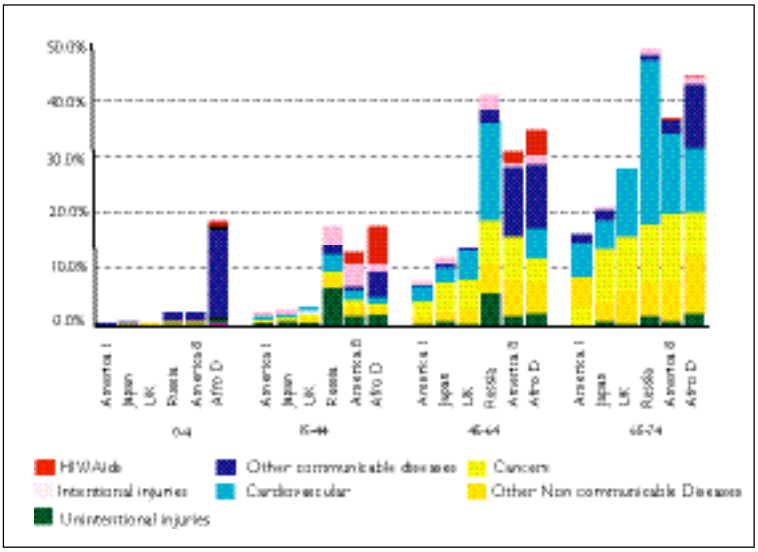
Health Inequalities

Inequalities in health is a vast topic, so I will not elaborate too much on it. I think the UK knows better than any country that the challenge of persistent inequalities in health needs to be addressed with a series of actions both inside and outside health systems.

Let me illustrate the underlying theme when we try to unpack inequalities at the macro level, looking at children, adult men and adult women. One illustration of health inequalities uses data from the US, Japan, UK, Russia, and West Africa. We divide the life cycle into four components and compute the probability of death between birth and age five, between 15 and 45, 45 and 65, and 65 to 75 for these countries. This is shown in Figure 7.

We have divided the US using local county-level data into eight groups. Figure 7 shows only two of these groups. America 1 refers to those Americans living in the counties with the highest life expectancy - the top two and a half per cent of Americans. America 8 refers to the bottom two and a half per cent of Americans.

Figure 7 **Probablilities of death in 4 periods of the lifespans, 8 Americas and the world: males, 1998**



The bottom two and a half percent of Americans have child death rates equal to Russia, much higher than the UK, and obviously many times higher than the US. However, in contrast to West Africa, there is an order of magnitude difference in child mortality. Child mortality is ten times lower in even the worst-off Americans than in West Africa. Remarkably for adult males the same is not true. Levels of adult male mortality in Russia, America 8, and West Africa are quite similar and are dramatically higher than the best-off Americans, the UK on average, or Japan. This phenomenon is present in middle-aged adults as well, with much higher levels and equal to West African levels in the broad sense as compared to the best-off Americans, as well as the average for Japan. The gradients are still present at older age, but in relative terms are clearly much smaller and in absolute terms only slightly smaller.

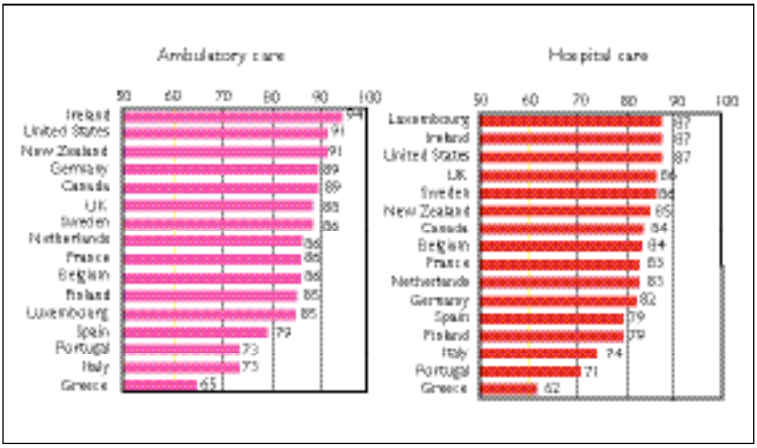
The message highlighted in the figure regarding inequalities related to adults, particularly adult men, is not a popular one in many political discussions of health policy. It is much easier to create a social consensus about tackling the problems of children, and in some cases of mothers, than it is to address the health problems of adult men.

Responsiveness

Responsiveness is an area where there has been remarkable progress on the measurement front, especially through the development of household survey instruments. I will use results for the UK to illustrate some of these developments. Based on the analysis of the anchoring vignettes for responsiveness included in the Multi-Country Survey Study, it appears that UK respondents are more likely to use favourable response categories such as "good" or "very good" for the same level of patient experience as compared to other European respondents.

Figure 8 shows the results for responsiveness when differences in cut-points are taken into account for a range of OECD countries included in this study. We look at a composite of ambulatory care - that is, adding up across the eight domains of responsiveness and the eight domains of responsiveness for hospital care. Some interesting findings are obtained.

Figure 8 **OECD health systems responsiveness**



First, the UK fairs rather well, even after taking into account the reduced expectations and the tendency to essentially get higher response categories. For ambulatory care the UK is number six on this list, behind Canada and ahead of Sweden; for hospital care it is placed

rather well as number four, and in absolute terms the gap between the UK, the US, Ireland, and Luxembourg is very small. The gap on outpatient or ambulatory care in absolute terms is larger between the top country (Ireland) and the UK, from 88 to 94.

Italy, Portugal, and Greece perform particularly poorly in both inpatient and outpatient responsiveness, which might be in line with people's expectations. A surprise to some might be Finland, doing quite poorly overall, and in particular in comparison to its neighbour, Sweden, which it often sees as its rival. Clearly there is a lot of detail in these responsiveness survey data.

One way of analysing that detail for a country is to look at the gaps in responsiveness. Rather than focusing on the achievement, we could consider what would it take, for the UK for example, to reach the highest level of responsiveness that is observed, and how much of that gap is attributable to the different domains? The diagnosis for inpatient and outpatient care is quite similar. Prompt attention - waiting lists, autonomy or involvement in decision-making, and communication from the provider to the patient comprise the majority of the issues in responsiveness for the UK. Perhaps not surprisingly, choice is a bigger issue for hospitals than for ambulatory care. Patient confidentiality in hospitals is more important than in outpatient settings.

Fairness in Financial Contribution

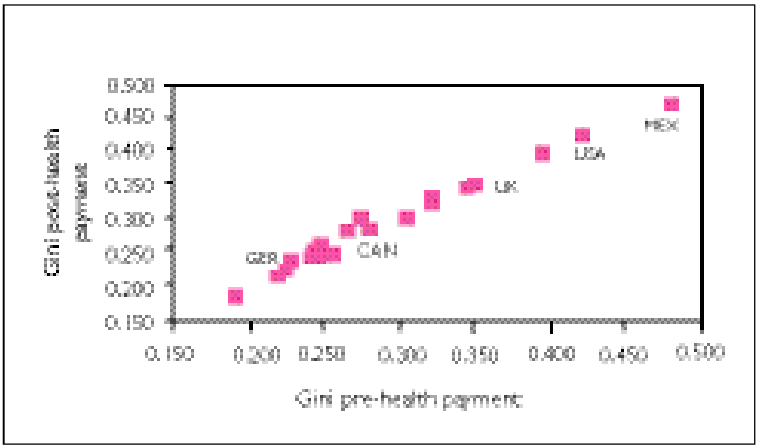
Since *The World Health Report 2000*, we have elaborated quite extensively our approach to analysing household contributions to the health system. There are four different approaches to this analysis, each providing different insights into health financing arrangements. We can ask the following question: "What is the effect of household health financing payments on household income?" The impact of health financing on income can be examined in terms of changes in the distribution of income or the fraction of households falling below the poverty line, due to health financing payments.

An alternative perspective is to recognize that the burden of a dollar on a household is much greater for a poor household than for a rich

one. We attempt to capture the burden which health financing payments inflict on a household by calculating them in terms of the percentage of a household's disposable income. For example, a 1,000 dollar payment by a poor household may represent 80% of the household's available income after paying for basic needs, and the same payment may represent only 1% of a rich household's disposable income. The burden on households can be quantified in terms of the fairness in financing coefficient or in terms of the percentage of households facing catastrophic health financing payments.

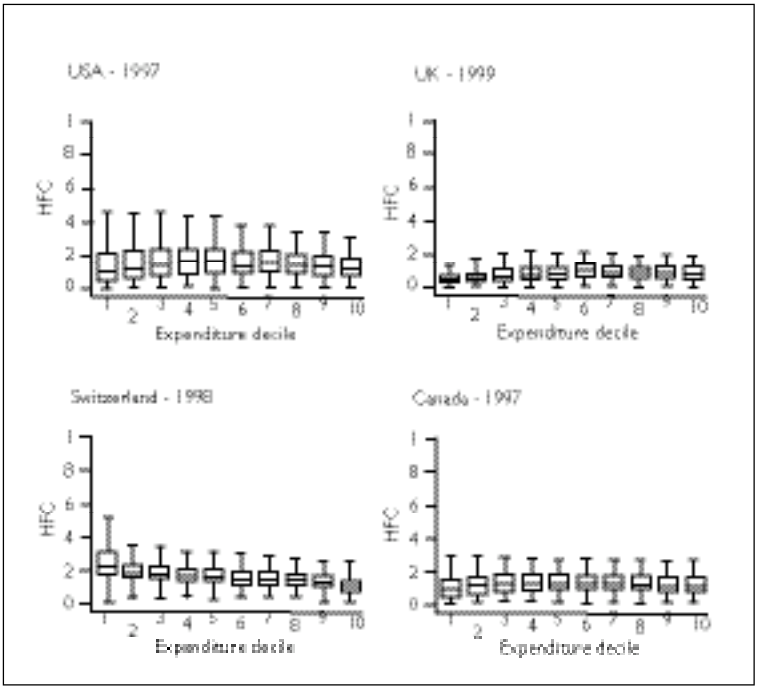
Figure 9 illustrates the impact of health financing contributions on the distribution of income. The x-axis shows the distribution of income in terms of the Gini coefficient prior to health payments and the y-axis shows the Gini-coefficient after such health payments. Payment here means out-of-pocket, insurance, social insurance, and taxes. First, the range in income inequality at the household level captured through surveys across OECD countries is staggering. Slovakia has the lowest Gini coefficient of 0.18, with Mexico near 0.5, and the UK at the higher end. If we showed this diagram in 1970, the UK would be closer down to Canada, but it has been drifting up towards the top right over the last 20 years.

Figure 9 **Health payments and income inequality, 20 OECD countries**



Second, health payments do not really make a big difference to income distribution. They seem to have little effect on the ordering of income distribution across these countries. They do have a marginal effect on income distribution, but it is not big. The variation across countries in income distribution is clearly enormous but not fundamentally related to health care payments.

Figure 10



If we look at the household level, we can focus on the burden that the US, UK, Switzerland, and Canada face for illustrative purposes. Using data from the household surveys for these four countries, we can show the range of the fraction of disposable income that households are paying for health care. For each country the ranges are divided into deciles, from the poorest (decile 1) to the richest (decile 10). In the US, for example, households up to the inter-quartile range pay 45 per cent of their disposable income for health care.

The UK performs remarkably well in this respect. There are very few

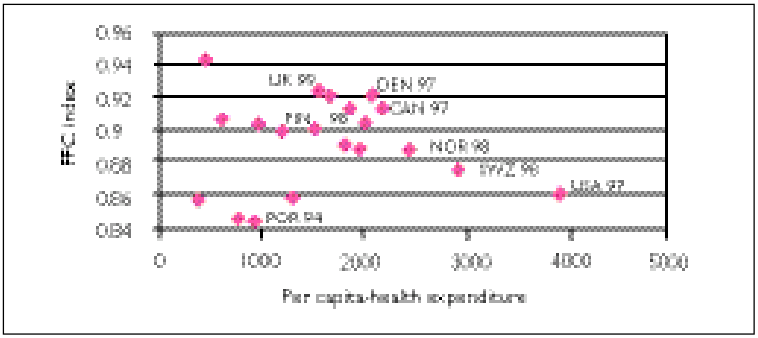
households in the UK spending more than 20 per cent of disposable income on health care. The poorest households are spending a lower share. It is rising up to about decile 4 and stays flat thereafter. In contrast, Switzerland, which has better income distribution than the UK but has a history of a mandatory flat payment for social insurance, has a very regressive pattern and wider variation in households' contributions to health systems at low income.

Canada's performance is not quite as good as the UK's. In Canada, there are older households that pay more than 20 per cent of their disposable income for health care. In the US, although it is progressive for the first four deciles because of Medicaid, there are tremendous problems of horizontal inequality or lack of financial risk protection - people without insurance or components of care which are not covered. At the same time, the UK is rather successful at protecting households from financial risk. So we can talk about two effects: both a regressivity problem and a lack of risk protection for the poorest households.

How much of this pattern is related to spending, given that the UK spends much less than the US and Canada? Is the good protection in the UK due to the instruments - social insurance or taxes - or due to the fact that the UK is spending somewhat less? Plotting an index of inequality, which we call the fairness in financial contribution (FFC) index, versus spending, summarizes those distributions. Figure 11 shows this relationship for different countries.

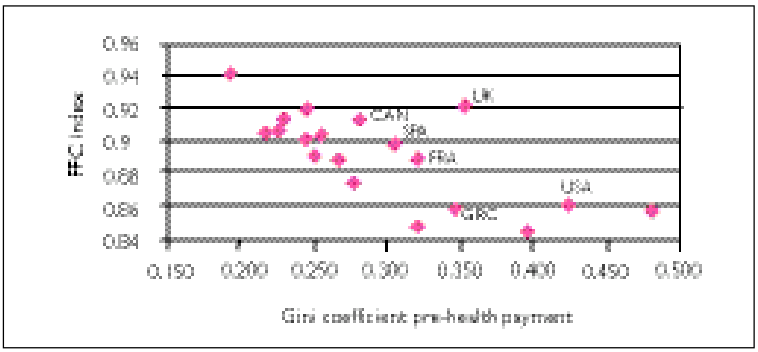
The UK in 1999 had the second highest level of the fairness index. The UK's index was equal to Sweden and Denmark's, ahead of Canada's and quite ahead of Finland's. We also observe that three countries stand out in the figure: Norway, which has quite a regressive tax structure for the health system; Switzerland, which we have already looked at; and the US, which has major deficits in risk protection and is at the high end of spending as well as the low end of the fairness index.

Figure 11 **Health expenditure per capita and fairness in financing, 20 OECD countries**



In Figure 12, the y-axis is our measure of fairness or household protection for heavy payments to the health system, and on the x-axis is income inequality. Countries above a Gini coefficient of about 0.3, excluding Canada, Spain, France, and the UK, have very unfairly financed systems and high failures of risk protection. There is then a gradient as countries are more equal. The same political economy that leads to equalization of income clearly tends to lead to more fairly financed health systems.

Figure 12 **Fairness in financing and income inequality**



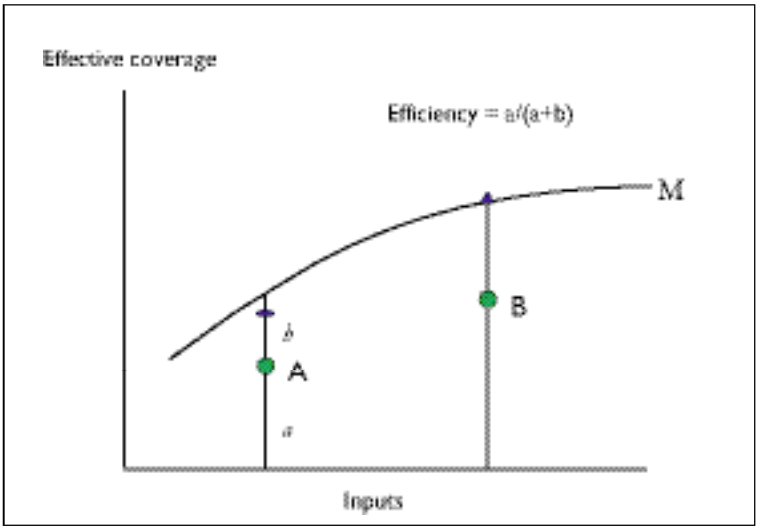
The exceptions are slightly off the line - Canada, Spain, and France - which have higher levels of financial risk protection or fairness of financing for a given level of income inequality. The spectacular

outlier is the UK where income inequality in 1970 was probably in line with Canada's current position. The UK preserved the financing system, despite the rise in income inequality. No other country seems to have done that.

Efficiency

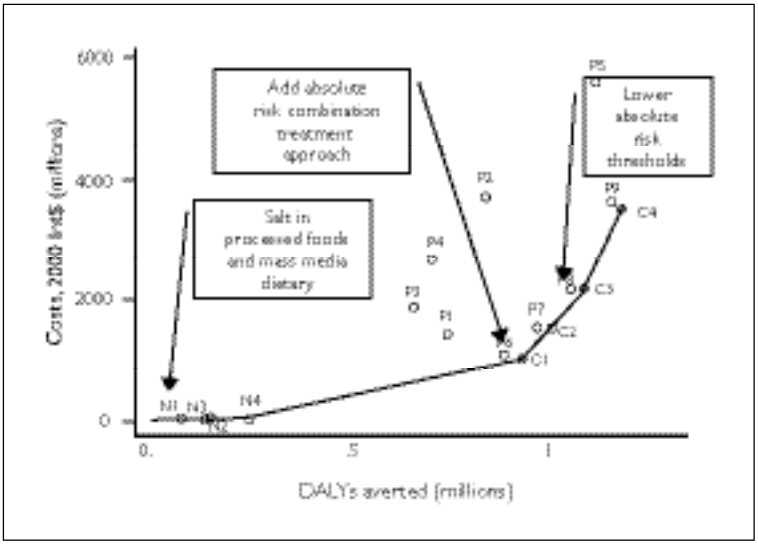
How do we get the frontier shown in Figure 13? Even if we have solved time lag issues and causal attribution problems by using information on intervention coverage, how do we figure out what is achievable for a given set of inputs?

Figure 13



One way, we believe, is through broader and more systematically available information on cost-effectiveness. Figure 14 is one illustration of that sort of expansion path or frontier; not overall, but just for one cluster of interventions. It is from our study on the cost-effectiveness of strategies to manage blood pressure and cholesterol. Each point is a combination of interventions. What it shows in absolute terms is the total cost of management and total health gain for cardiovascular or blood pressure and cholesterol strategies.

Figure 14 **Cost and effectiveness of selected CVD risk factor interventions, EURO A**

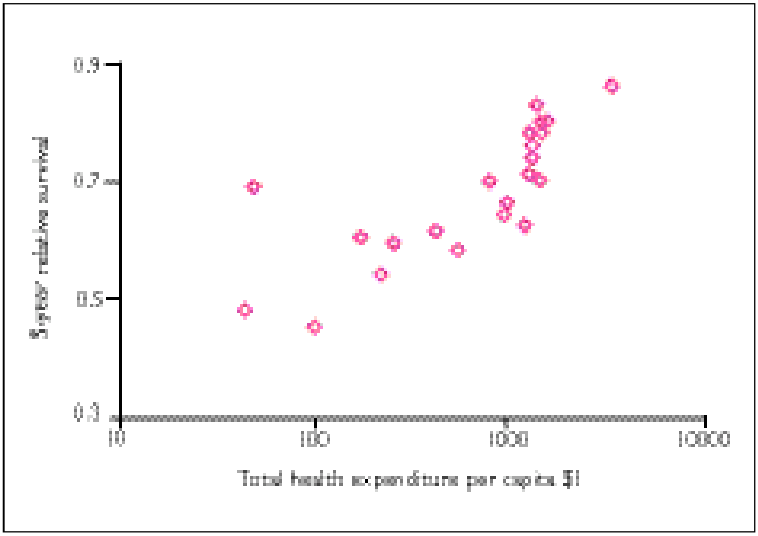


The bold line defines what we call the expansion path, i.e. what is the most attractive series of options to purchase as your budget increases? In the case of cardiovascular interventions, you would start by buying salt reduction, either through legislation or voluntary regulation and mass media programmes. Then, you would quickly move to adding in personal risk management using what is called the Absolute Risk Approach, that is, managing individuals in the clinical setting, their blood pressure, cholesterol, and adding an aspirin into the mix simultaneously. If you cross an absolute risk threshold for cardiovascular disease - stroke or heart attack, then you trigger a triple drug therapy, a statin, an anti-hypertensive, and aspirin. Those, both in the studies of impact and in cost-effectiveness, are extraordinarily attractive and have potentially huge health benefits.

We believe that we will be able to have this type of information for many interventions and we can try and construct the frontier to assess the system's efficiency. Although detailed effective coverage information for a range of interventions is not yet available for the UK or other OECD countries, Figure 15 explores the relationship between

spending and effective coverage using breast cancer survival data as a crude proxy.

Figure 15 **Breat cancer 5-year relative survival vs. health xpenditures**



Using data both from the Euro Care Study and other population-based cancer registries that are monitored at WHO, there is a strong relationship between spending and effective coverage. There is also quite a substantial variation at given levels of spending. The UK does not perform that well in this metric of effective coverage, even compared to countries at similar levels of spending. Depending on the interpretation, we can focus on the underlying production function, more dollars lead to increased coverage or the variation in the efficiency of resource use. Resource levels versus efficiency of resource use are key aspects of the health care debate in the UK.

UK Considerations

Let me provide some tentative lessons for the UK based on a very distant view from Geneva. Such a macro view benefits from distance from the policy debates, but is also severely hindered by a lack of detailed knowledge about the UK system and ongoing debates.

First, despite high levels of income inequality, the UK performance in protecting particularly poor households from financial risk is quite remarkable. Second, from a public health perspective, there does seem to be room for improvement of health outcomes and health systems relative to other OECD countries, not only through more activist roles in prevention (not simply making services available, for example, on individual risk management), but also by figuring out ways to encourage people to use them. Given the breast cancer survival data and other indications, the UK can narrow the gap in health outcomes by increasing the gap in health outcomes by increasing the quality of care. This is not a criticism of providers. The root cause of low effective coverage may simply be a lack of resources.

Responsiveness in the UK system is extremely high, given spending levels, but the data seem to suggest that there is room to expand on issues around dignity, involvement in decision-making, and communication between providers and patients. It is our belief that investment in health information systems which include effective coverage would provide a powerful tool, particularly in a country with as sophisticated decision-making community and public as the UK, in directing policy improvements in the future.

Final Reflections

Let me close with a few reflections. There has been a lot of debate and criticism about *The World Health Report 2000* and our work on performance assessment. In addition to the debate we have seen some interesting ramifications of that work. Four are worth highlighting:

1. First, in a number of countries where financial risk protection or catastrophic expenditure or impoverishment due to health payments was not on the political agenda, it is now firmly there. Mexico, for example, just passed its most substantial reform of the health system in 60 years extending insurance coverage to the entire population, including 50 million new individuals, due to the issues raised around financial risk protection in *The World Health Report 2000*. Similar events are happening in Iran. No reform has yet been passed in China but

discussion along those lines is ongoing.

2. A second type of comment we have received from a number of governments is that by this work, and by saying that there are multiple outcomes of health systems, we have legitimized to some extent the construct of patient experience or interpersonal quality of care or responsiveness as an outcome that is important to the population.
3. In a number of developing countries, there has been a small, but important shift from an exclusive focus on vertical disease-specific programmes to more discussion of health system issues.
4. In countries which previously had shown no interest in data systems, there have been serious attempts to improve national health information systems.

After five years, we still believe that cross-national research on health systems can and will provide insights into what works and what does not. These insights will only come if many of the nations of the world use a common framework, definitions and metrics, for inputs, processes, coverage and outcomes. This investment in health information or health metrics is not something that has a quick pay-off; rather it must be seen as a long term enterprise.

I think the lesson that I take from all the discussion is slightly different. The most important message to me is the notion that by providing certain types of information we can foster a culture of accountability for outputs and outcomes. This may be the most significant ingredient, at least from the international perspective for improving health systems. Many of the challenges and difficulties that systems face are very local, so there are no global solutions or recommendations for addressing them. On the other hand, if there is a culture of accountability and accountability mechanisms create positive incentives for improving performance, local entrepreneurs, or social entrepreneurs as some people call them, will find solutions.

To have that type of culture of accountability, however, requires:

1. credibility of the measurement;
2. clarity of the measurement; and
3. comparability.

Measurement must be credible, both in reality and in the perception of everybody involved. Credibility requires: valid and reliable measurements, and a complete data audit trail for information that is used in performance monitoring. The explicit data audit trail must include putting in the public domain. Microdata, all methods and approaches used to correct for known bias, analytical models, and any procedures used for missing data imputation. Such an explicit data audit trail is good common sense but has not been widely implemented. This administration of WHO is committed to this concept, although it will take quite a while to achieve.

The second aspect of accountability is clarity - clarity of communication does not mean simplicity. Clarity requires that what is being measured is easily understood by the media and decision-makers, and is quantified in some meaningful units. Academics and technical specialists will always need further disaggregation of information to diagnose problems and test hypotheses. Such disaggregated information is essential, but to foster accountability there must also be clarity of what is being measured and what it means to a broad audience.

For accountability you need some simple measurements that can be understood and defended, that are credible and clear to people. For example, I think impoverishment due to health system payment is clear - complicated to measure, but very clear. Healthy life expectancy is also rather clear. Effective coverage can be quite clear as well, and our experience with both heads of state and ministers of health suggests that this is true.

Lastly, there is the notion, very unpopular with some, that accountability actually requires comparability of measurement. If you have measures that are only locally-specific and cannot be compared with anybody else, then you decrease the scope of accountability enormously. Imagine that you are looking at air conditioners and you are going to see if they work, but you are using a thermometer that only has a scale unique to that thermometer. Yes, you can see if the air conditioner reduced the room temperature, but you will not be able to make any comparisons to anybody else. It is as you broaden the comparisons that the relevance and the scope of accountability expand.

Most of the ideas, methods, and results presented in this lecture are more fully explored in the WHO volume, *Health Systems Performance*

Recent OHE publications

To Publish or Not? Experience and Evidence about Publishing Hospital Outcomes Data
A Mason, A Street, 2005 (price £12.50)

Activity Based Funding for Hospitals: English Policy, International Experience
By John Sussex and Andy Street, 2004 (price £12.50)

What Could be Nicer than NICE?
By Alan Williams, 2004 (price £7.50)

The Impact of Ageing on Health Care Expenditures - Impending Crisis, or Misguided Concern?
By Meena Seshamani, 2004 (price £12.50)

How Much Should We Spend on the NHS? - Issues and Challenges Arising from the Wanless Review of Future Health Care Spending
Edited by John Appleby, Nancy Devlin and Diane Dawson, 2004
((price £12.50)

The Economics of Dental Care
By Ray Robinson, Darshan Patel and Rowena Pennycate, 2004 (price £12.50)

Developments in Economic Evaluation in Health Care: a Review of HEED. OHE Briefing No.40
By Clive Pritchard, 2004 (price 7.50)

Mental Health Economics and Policy in a Global Context. Seminar Briefing No.5 Edited by Jon Sussex, 2003
(price £7.50 for hard copy, downloadable free from www.ohe.org)

Reducing Harm to Patients in the National Health Service. Will the Government's Compensation Proposals Help? OHE Briefing No.39
By Adrian Towse, Paul Fenn, Alastair Gray, Neil Rickman and Rodrigo Salinas, 2003 (price £7.50)

To Heal and Harm: an Economic View of Drug Safety
By Jonathan Silcock and Clive Pritchard, 2003 (price £12.50)

Details of all OHE publications, and how to order them, can be found on the OHE website at **www.ohe.org** or by contacting:

The, Office of Health Economics, 12 Whitehall, London, SW1A 2DY
Tel: +44 (0)20 7930 9203, Fax: +44 (0)20 7747 1412