
Chapter 15.1

SUMMARY MEASURES OF POPULATION HEALTH: CONCLUSIONS AND RECOMMENDATIONS

CHRISTOPHER J.L. MURRAY, JOSHUA A. SALOMON,
COLIN D. MATHERS AND ALAN D. LOPEZ

INTRODUCTION

The regular assessment of population health is a key component of the public policy process to improve health levels and reduce health inequalities in the World Health Organization (WHO) Member States. Population-level estimates of mortality, morbidity and health states in countries by age, sex and cause, are useful for numerous public health purposes, ranging from the monitoring of new epidemics to measuring progress in reducing old ones for which disease control programmes are in place. To describe health patterns adequately in almost 200 Member States according to age, sex and cause, a vast array of estimates need to be generated. It then becomes a major challenge to interpret the key findings of such a review or to compare levels of population health across countries unless the data are summarized in some fashion.

The simplest and most widely used method for producing population health statistics is to aggregate data on individuals in order to generate statistics such as the proportion of the population (or of a particular population subgroup) suffering from a given health problem or living in a particular health state, or the number of individuals who die from a particular cause during a specified interval. This approach rapidly becomes unwieldy when a number of problems are being monitored and we want to make comparisons over time, across population groups, or before and after some health intervention.

Summary measures of population health (SMPH) are measures that combine information on mortality and non-fatal health outcomes to represent population health in a single number. While such summary measures have many potential uses (see part 2), there are two that are particularly important for public health policy: (1) comparisons of the average health levels in different populations or subgroups, or in the same population over time; and (2) assessments of the relative contributions of

different diseases, injuries and risk factors to overall population health. These two key uses may best be served by different forms of SMPH.

For reporting on the health of WHO Member States, SMPH provide a convenient and useful summary of the vast array of components of population health. SMPH do not replace the more detailed reporting of data on specific aspects of health and mortality or on the specific causes of health problems; rather they supplement these data with more comprehensive indicators that can be used to monitor trends and compare levels of health across populations. During the 1980s and 1990s, and particularly since the publication of the first results of the Global Burden of Disease Study in the 1993 *World Development Report*, there has been a growing debate on numerous aspects relating to the development and application of summary measures of population health. We hope that this book will provide a major contribution to that debate by assembling the views and arguments of health policy-makers and experts from a wide range of disciplines including epidemiology, demography, health statistics, health economics, philosophy and ethics. In this concluding chapter, we summarize the important conceptual, empirical and ethical issues identified and debated by contributors, and draw some conclusions and recommendations for the future evolution of summary measures of population health.

CONCEPTS

DEFINITION OF HEALTH

The conceptual basis for summary measures of population health begins, logically, with the definition of health. The Constitution of the World Health Organization famously defines health as “a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity”, and notes that the “health of all peoples is fundamental to the attainment of peace and security” (WHO 1948). In the broadest terms, this definition sets forth a lofty ideal for health as an integral component of well-being and, further, expresses the notion that good health is a necessary condition for attaining the highest possible levels on all other aspects of well-being. In defining health in terms of an ideal, the WHO Constitution provides the first building block for an operational definition of health. Over the half-century since this definition was set forth, there have been continuing efforts to develop more precise conceptualizations of health that may be linked to operational measures.

From these efforts, one common theme that has emerged is that health may be viewed naturally as an intrinsic, multi-dimensional attribute of individuals; there is an intuitive understanding of health that crosses cultural boundaries, such that when we talk about a person's health, we are understood to be referring to his or her levels on the various components or domains of health. It is important to note that this view

distinguishes health states from pathologies, risk factors or etiologies, and from encounters with health services or the application of health interventions. This conceptualization preserves the spirit of the WHO Constitution definition: rather than equating health with diseases or diagnostic categories, it recognizes a causal chain through which risk factors are determinants of diseases, and diseases in turn are determinants of health states.

A formal framework for cataloguing the multiple domains of health has been developed by WHO in the International Classification of Functioning, Disability and Health (ICF; WHO 2001a; also chapter 7.3). The ICF replaces the concepts of *disability* and *handicap* in the International Classification of Impairment, Disability and Handicap (ICIDH) with the concepts of *capacity* and *performance*. *Capacity* refers to an individual's ability on a domain as it would be manifested in a uniform environment (assuming motivation to perform the task), for example, the ability to walk 100 metres on a level, well-lit, non-slippery surface. *Performance* describes an individual's ability on a domain as it is manifested in his or her current environment. The gap between capacity and performance therefore reflects the impact of an individual's actual environment (and perhaps motivation) relative to the uniform environment. Both performance and capacity may be measured either with or without an individual's personal aids.

Given this distinction between capacity and performance, which construct do we aim to capture in summary measures of population health? To the extent that performance reflects an individual's unique environmental setting, which may also vary with time and as individual circumstances change, it is probably not congruent with most notions of health. We would not want to characterize the same cognitive impairment differently in two individuals simply because they have different vocations that call upon different types of cognitive tasks. Similarly, we would not say that an individual with a hearing impairment is healthier simply because she avoids noisy gatherings. These examples point to a common-sense understanding of health that does not correspond to performance because it excludes the idiosyncracies of an individual's environment.

The notion of capacity corresponds more closely to this common-sense interpretation of health by defining external environmental factors in a uniform way. More specifically, we believe that capacity *with* an individual's currently available personal aids is the most appropriate construct, as many societies commonly understand personal aids such as crutches, glasses or hearing aids to improve health levels on relevant domains. There remains a degree of arbitrariness in defining the boundary between personal aids and individual-specific environmental factors, but we believe that the inclusion of simple personal aids that should in principle be available to all people is most consistent with the intuitive societal notions of what constitutes health.

KEY DOMAINS OF HEALTH

In conceptualizing health as a multi-dimensional construct, we recognize that health domains constitute a subset of all of the domains of well-being. Having postulated this conceptual boundary, there are likely some core domains of health that almost all people would agree upon, while other domains might be less universally accepted as health domains. We may also identify domains of well-being that are not strictly health domains but can serve as good proximate measures of the experience of health (*health-related* domains). For operational purposes, there are some instances in which the best or only measurable indicators of the levels on certain health domains may in fact be consequences that are outside the realm of health.

In order to operationalize the measurement of health, we need to define a parsimonious set of domains that capture the components of health that are generally considered most important to people in determining their overall assessments of health levels. Although domains of health may overlap, the aim is to identify a set of domains that is exhaustive as well as generally acceptable as the content of the ordinary meaning of health. Chapter 7.1 proposes that key domains of health should be:

1. valid in terms of intuitive, clinical and epidemiological concepts of health;
2. linked to the conceptual framework of the ICF;
3. amenable to self report, observation or measurements;
4. comprehensive enough to capture all important aspects of health states that people value; and
5. cross-population comparable.

Although existing health state measurement instruments have differed considerably in their content in an attempt to arrive at a set of domains that covers the universe of health adequately, there are several domains that have been included in nearly all generic measures of health states, including pain, affect, cognition, mobility, self-care, usual activities (including household and work-related activities) and interpersonal relationships. Among these, it may be argued that some are core domains of health (e.g. mobility, pain, cognition) while others may be considered as proxy measures of multiple domains (e.g. self-care, usual activities).

THE BOUNDARIES OF GOOD HEALTH

Another issue that needs to be addressed in operationalizing a definition of health is whether all increments and decrements on a domain are understood as improvements and losses of health, respectively, or whether there is some threshold above which increments and decrements are not perceived as changes in a person's health state. For example, should one consider a person with an IQ of 180 as being healthier than another individual with an IQ of 150? Or should one say that the former is not

necessarily healthier by virtue of a capacity that exceeds some norm for cognitive excellence? We believe that the concept of a threshold for full health accords better with commonly held societal views of health than an allowance for unbounded improvements in domain ability to be considered as improvements in health. The “supra-health” levels are perhaps better referred to as talent.

Some have argued that the threshold for full health can be identified purely in biological terms by examining the statistical distribution of functioning in the domain (Boorse 1977), but it seems clear to us that the domain threshold for full health is a normative choice: there is no criterion that would allow us, *a priori*, to choose a particular point on the population distribution of domain ability as representing the threshold for full health. We therefore suggest that the identification of thresholds for domain ability should be empirically-based and linked to health state valuations (described below). In intuitive terms, the threshold for a particular domain is the level of ability below which a majority of people generally recognize decrements as departures from excellent health. In practice, average health state valuations for levels of ability close to or above the threshold level will be extremely close to the value for full health, so that it is not necessary to explicitly delineate each domain threshold.

HEALTH STATE VALUATIONS

Thus far, we have elaborated a conceptual framework for health, in which health states are described in terms of levels on multiple dimensions such as mobility, pain, hearing and seeing. Health state valuations, in relation to these multidimensional profiles, constitute scalar index values for the overall levels of health associated with different states, measured on a cardinal scale that ranges from zero (for a state equivalent to death) to unity (for a state of ideal health). These valuations formalize the intuitive notions that health levels lie on a continuum and that we may characterize an individual as being more or less healthy than another at a particular moment in time. Health state valuations quantify departures from perfect health, i.e. the reductions in health associated with particular health states. It is important to emphasize that these weights *do not* measure the quality of life of people with disabilities and *do not* measure the value of different people to society.

In fact, there have been a variety of different conceptual interpretations of health state valuations that have led to considerable confusion in defining the basis for measuring and understanding these valuations. It is useful for us to contrast our conceptual definition of health state valuations with these other concepts.

Many health economists have defined health state valuations as measures of the utility associated with health states (see chapter 10.2). Our notion of health differs from utility in two important respects. Firstly, utility represents a broader construct, corresponding more closely to well-being than to health, strictly defined. Further, utility in the health context

is typically defined in terms of the von Neumann-Morgenstern axioms of expected utility—i.e. conceptualized on the basis of choice under uncertainty. The use of the standard gamble technique for elicitation of valuations is linked to the expected utility framework, and therefore invokes both assessments of health levels associated with different states as well as attitudes towards risk and uncertainty. Thus, the notion of utility in the context of health state valuations combines our concept of health with the separate concept of risk aversion, which we do not believe is relevant for characterizations of health levels in summary measures (see chapter 9.4).

Another common interpretation of health state valuations has been as measures of quality of life (QoL), a term that has been used widely in various social science contexts to refer to the overall, subjective appraisals of happiness or satisfaction experienced by individuals. In health, the term QoL has been used often in a more particular way to refer to a multi-dimensional construct relating to symptoms, impairments, functional status, emotional states and what we have labeled as health domains. This use of QoL is clearly inconsistent with the general use of the term, so health researchers have taken to referring to this construct as “health-related QoL.” To the extent that an individual’s health-related QoL is conceived of as a vector of levels on “health-related” dimensions of life, it is similar to our conceptual framework for measuring health, albeit with less precisely articulated boundaries. Where health-related QoL is viewed as a summary measure of the contribution of an individual’s health to her overall well-being, on the other hand, conceptual problems emerge from the fact that well-being is not clearly separable into independent health and non-health components, as Broome has argued convincingly (see chapter 3.1). In other words, when we compare the well-being or “quality of life” of individuals with different health levels, these relative comparisons may change depending on their levels on non-health dimensions of well-being. It is in this sense that a person with a long-term disability may say that their health-related quality of life is better than that of another person with no long-term disabilities.

We may avoid these difficulties if we define health state valuations to be simply indices of overall levels of health. Unlike the notion of utility, we do not believe that it is necessary to define this construct explicitly in terms of choices or preferences. Almost everyone can agree that a person with one amputated leg is healthier than a person with two amputated legs, all else being equal, without resorting at all to either the language of choice or to statements about the overall well-being of either person. While this is a simple case of a dominance ordering (because the difference is in the level of only one domain), the same intuitive notions apply to more complicated examples: if we say that somebody with a mild sore throat is, *ceteris paribus*, healthier than somebody with two broken arms, perhaps not everyone would agree, but most people could at least understand our statement through some common-sense notion of health. Indeed, this

common-sense notion extends beyond ordinal comparisons, for example, allowing us to say that going from good health to having a sore throat is a smaller change in health than going from a sore throat to quadriplegia. In all of these cases, we submit that there is an intuitive understanding of the meaning of health that is not based on the concept of choice. Sen (1992) has similarly argued that welfare values do not need to be choice-based.

It is important to emphasize that, in defining health without reference to choice or preferences, we also distinguish our notion of health from Broome's broader notion of the "goodness" of health (chapter 3.1). In so doing, we avoid the problem of the separability of health and non-health components of well-being. Broome is undoubtedly correct to question whether the "goodness" of health—whether it is for comparative judgement or for some ultimate policy use—is truly separable from the "goodness" of non-health well-being. As Hausman argues in chapter 3.2, however, almost every culture has some notion of health that is distinguished from overall well-being, and it is this intuitive notion of health that we aim to capture in summary measures of population health. The World Health Organization is fundamentally concerned with the improvement of health as distinct from aspects of well-being that are not intrinsic determinants of health, and therefore requires an appropriately defined measure of population health.

AGGREGATION ACROSS TIME AND INDIVIDUALS IN POPULATION HEALTH MEASURES

In chapter 3.1, Broome raises some fundamental issues about the aggregation of health (more specifically the goodness associated with health) across time and across individuals. As we have emphasized here, our notion of health focuses more directly on the levels on a set of domains, rather than on the goodness associated with health, or overall well-being or utility. This focus, together with our careful consideration in chapter 1.3 of the different meanings of the term "healthier than" at the individual level, allows us to address the issue of aggregation in a fairly straightforward way.

Moving from the measurement and valuation of health at a moment in time, chapter 1.3 describes two other key perspectives that may be used in answering the question "Is person A healthier than person B?" The first is to compare the totality of health states experienced by person A and person B over their entire lifetimes. The second is to consider only the current health and future prospects of the two individuals. The latter perspective is probably closest to a common-sense notion of whether one individual is healthier than another. In this view, the past is excluded, but the influence of the past on current or future health is captured. The comparison of individual health in this perspective is based on each individual's health expectancy—the expectation of years of healthy life.

Aggregation across individuals presents another set of challenges, and different approaches may be required depending on the intended use. For

example, health expectancies rely on a life table approach in which a hypothetical birth cohort is exposed to population rates of mortality and non-fatal health outcomes. In this case, there is no real aggregation across any real set of individuals. We may, however, construct a measure of the average health expectancy in the population which does result from an aggregation of the currently living members of the population (see below). For health gaps, a range of approaches are possible (see chapter 5.1), in which different sets of individuals and different time horizons are used as the basis for aggregation.

One important question is whether to include distributional concerns directly in summary measures of population health as part of the aggregation rule, or whether, on the other hand, to produce separate measures of average levels of health and distribution of health. Chapters 3.1, 3.8 and 3.9 address various viewpoints on this issue, and we discuss it further below.

HEALTH EXPECTANCIES

As summary measures of the overall level of health of a population, health expectancies have two advantages over other summary measures. The first is that it is relatively easy to explain the concept of an equivalent “healthy” life expectancy to a non-technical audience. The second is that health expectancies are measured in units (expected years of life) that are meaningful to and within the common experience of non-technical audiences (unlike other indicators such as mortality rates or incidence rates).

We can categorize health expectancies into two main classes: those that use dichotomous health state weights and those that use health state valuations for an exhaustive set of health states. Disability-free life expectancy (DFLE) is an example of the first class (Robine et al. 2001). Healthy life expectancy (or HALE) is an example of the second class (Mathers et al. 2001b; WHO 2001b). Health state expectancies such as DFLE give an implicit value of zero (equivalent to the valuation of death) for disability above a certain threshold; below this threshold the valuation is 1. This means that the summary indicator is not sensitive to changes in the severity distribution of disability within a population (criterion 5 of Murray et al.—see chapter 1.2). Additionally, the overall DFLE value for a population is largely determined by the prevalence of the milder levels of disability, and comparability between populations or over time is highly sensitive to the performance of the disability instrument in classifying people around the threshold. For these reasons, health state expectancies are not appropriate for use as SMPH.

In contrast, HALE is sensitive to changes in the distribution of health states, and thus meets one of the key criteria for an acceptable SMPH. Although healthy life expectancy cannot be additively decomposed with respect to causes or population subgroups, it is additively decomposable into health expectancies for specified levels of disability severity. Health state expectancies should thus be better understood as decompositions of

summary measures than as SMPH in themselves. This interpretation is consistent with the usual ways in which families of health state expectancies are presented for a population (see chapter 4.1).

The calculation of health expectancies requires explicit choices about the time perspective adopted for defining event rates for the various input parameters, namely incidence, remission, prevalence and mortality. Thus, for example, one could calculate health expectancies as a pure *cohort* measure, where all events are based on observed events in a cohort. This would require tracing a birth cohort until every member had died. Alternatively, cohort event rates by age could be estimated based on projections.

Health expectancies could also be pure *period* measures, analogous to the standard demographic approach to calculation of life expectancies, where all event rates are calculated for a hypothetical cohort which is assumed to experience current age-specific event rates throughout life. These event rates would be the complete set of age-sex specific transition rates between all health states (including death).

The three perspectives identified in chapter 1.3 (health at a moment in time, health over the entire lifespan, or current and future health) provide a basis for constructing aggregate health expectancies across time and cohorts for populations. In principle, the population health expectancy defined as the average of all the *future* individual health expectancies for the people comprising the population at a given point of time satisfies all five criteria for SMPH (see chapters 1.2 and 4.1). This is because the future individual health expectancies are dependent not only on current and future transition rates (incidence) but also on the current health status (prevalence) of each individual.

Future individual health expectancies require assumptions about future incidence, remission and mortality rates that the individuals will face, whereas currently computed period health expectancies provide a measure based only on currently measurable aspects of health. For this reason, we may choose to define and estimate individual health expectancies by assuming that future incidence, remission and mortality rates that the individuals will face reflect current health conditions only. This would be an analogue of the well-known period life expectancies and health expectancies. We can then calculate a population-based aggregate health expectancy that satisfies the five criteria. For some comparisons, the average health expectancy could be based on a standard population to remove the effect of differences in age structure between populations.

The simplest method is to base the calculation of healthy life expectancies on the currently observed period information—particularly mortality rates and health state prevalences. This allows us to compute health expectancies for populations using Sullivan's method (see chapter 4.1). This approach requires only the data required for the period life table together with population prevalences for health states which can be

measured using cross sectional population surveys and/or burden of disease analyses for prevalent disability in populations.

If sufficient data were available, it would be possible to use a multi-state life table based on the current observed period transition rates between all health states (and death), but to adjust these transition rates to reflect current risk exposures. Thus for example, the risks of lung cancer at each age could be based on current patterns of exposure to tobacco smoking, rather than on currently observed incidence or mortality of lung cancer (which reflects exposure 20 years or more in the past). This would be an extension of the concept of a period measure to include current risk exposures. Note that whereas a true period measure must be based on incidence rates (transition rates) for health states, it should reflect prevalence (rather than incidence) of relevant risk exposures, since it is the prevalence and duration of exposure that determines risk of transition between health states.

HEALTH GAPS

The second main use of summary measures of population health is to assess the relative magnitude of the contribution of different diseases, injuries and risk factors to levels of population health. Such information is a useful input to debates on priorities for research and development, priorities for focused attention of government for policy formulation and for identifying which interventions should be further evaluated. When using summary measures to look at different causes of population health levels, the critical question is different: Is cause A or B a more important contributor to levels of population health? To be comprehensible for the broader public, a summary measure used for such causal attribution should fulfill two important requirements: it should be easily understood and it should have the property of additive decomposition. In other words, the summary measure should be partitionable into causes or subgroups such that the sum of the SMPH across a set of mutually exclusive and exhaustive categories equals the total.

In general, health gaps can be decomposed into the contribution of various causes in a more intuitive and easily communicated fashion than health expectancies. DALYs are additive across causes to give the total health gap for a population. A health gap measure such as the DALY thus fulfills different needs for SMPH to be used for causal attribution. Part 6 of this book presents a series of viewpoints on the two possible approaches to causal attribution: categorical attribution and counterfactual analysis.

Since the primary purpose of measuring health gaps is to disaggregate them into the contributions of component causes, they should be formulated to capture the relative magnitude of causes of population health currently and in the near future. The simplest method is to base the calculation of health gaps on the currently observed period information—particularly mortality rates and health state prevalences and transition rates. This allows us to compute health gaps for populations using only

period data for mortality and health states in populations. The most widely used implementation of this approach is DALYs.

The key conceptual issue relating to health gaps is the choice of the normative goal for health against which current conditions are compared. Chapter 5.1 discusses three different types of normative functions that may be defined: a population norm, an unconditional individual norm and a conditional individual norm. A population norm simply specifies a reference healthy life expectancy, allowing the calculation of a population health gap as the difference between the realized healthy life expectancy and this reference. Like healthy life expectancy, this type of health gap is not easily decomposed into the contributions of different causes. Furthermore, because it represents a life-table type measure aggregated across ages, it does not distinguish between different age patterns of health levels that produce the same overall health expectancy, even though norms for health may vary considerably by age.

An unconditional individual norm defines a set of age-specific target levels for health that do not change depending on the age an individual has attained, while a conditional individual norm may vary as individuals advance in age. For example, an unconditional norm may be defined as survivorship in full health up to age 100. An individual who lives to 101 in full health thus exceeds this norm and contributes nothing to the health gap measure. In contrast, a conditional norm, such as the two-part norm implemented in the DALY, may set a survivorship goal that is specific to the age that has already been attained. In this case, an individual who lives to age 101 in full health may still contribute to the health gap, given a norm for healthy life expectancy at age 101 that is greater than 0. The conditional norm allows for all individuals in the population to contribute to the total health gap, no matter what age they attain.

Criteria for selection of a specific normative loss function might include biological plausibility (in principle, the normative goal should be biologically achievable), desirability and, perhaps, some notion of fairness (for example, the fair innings concept described in Williams 1999). Based on the observed experience across countries, a reasonable normative survivorship function to select as the basis for defining a standard loss function would be one that equals the highest national life expectancy observed across populations, at present a life expectancy at birth of around 85 years. Loss functions based on local survival curves should not be used for international comparative health gap analyses, although such loss functions may be used for various other types of national studies.

ETHICS

THE ROLE OF EMPIRICAL ETHICS

In the conference behind this book and in subsequent discussions, one of the most lively debates has revolved around the question of empirical ethics

(see part 12). Some authors have argued that any values that are incorporated, either implicitly or explicitly, in summary measures should be based solely on the empirical measurement of these values in the public, and that the entire enterprise of constructing summary measures is threatened by heterogeneity in values. For example, Mooney states in chapter 12.3:

If the two (or more) populations are genuinely comparable in terms of their value bases, then comparing their health status might well be justified. The occasions on which such genuineness will be present, however, seem rather few and the larger the population groupings involved, the less likely that genuine homogeneity will apply.

The extreme version of this argument would suggest that all comparisons are impossible because of differences in individual values. Despite assertion by proponents of empirical ethics that there is extreme heterogeneity of values, empirical evidence suggests otherwise, at least in relation to health. Valuation studies carried out with deliberative small groups in 12 different countries have found surprising consistency in valuations across cultures (Murray and Lopez 2000). More recently, valuation studies carried out as part of the WHO Multi-Country Survey Study (Üstün et al. 2001) also have found remarkable consistency in health state valuations. This result is not, perhaps, a major surprise, given the fundamental importance of the core health domains to all human beings, irrespective of their social or socioeconomic circumstances. It is possible, though again the empirical evidence is not apparent, that there may be more heterogeneity in values around well-being outcomes than health outcomes, conceptualized as described above, as a set of domain abilities. Given the fundamental importance of the core health domains for the achievement of the things that people commonly value, we believe that comparison of health levels through SMPH is not only meaningful, but operationally achievable.

In fact, even if there is heterogeneity of values, other disciplines have found very straightforward ways to address this issue. For example, consumer price indices are computed in reference to a standard basket of goods, even though individual consumption patterns may vary widely. In descriptive epidemiology, where the age structures of populations differ, we use one common standard in order to make comparisons. This does not conflict with use of local standards for other purposes. Similarly, even though there may be variation both within and across countries on values that are incorporated in SMPH, for purposes of international comparison, it is quite appropriate to use a common standard.

Hausman in chapter 12.2 presents a compelling critique of empirical ethics, concluding as follows:

Although there is a great deal to be said for finding out what people value, there is little to be said for empirical ethics... The main problem is that the defender of empirical ethics poses a false dichotomy:

either statisticians and health professionals decide, or the values of the target population rule... [T]here should instead be a dialogue between the values of the target population and the moral arguments of those who have laboured long over these issues. If the moral arguments of the statistician or health analyst are conclusive, then they should transform the social consensus. If there is wisdom in the social consensus (and the experts are sensitive and open-minded enough to recognize it), then the values of the population will reveal narrowness or flaws in the arguments of the experts. If the values within a society are genuinely moral values, then they must be both sensitive to rational argument and rationally defensible. Nothing should prevent the analyst from influencing the population's values and being influenced by them. Those who study health should have no authority beyond that conveyed by their arguments. But on moral issues those concerned with health should defer only to arguments, not to mere consensus. Empirical ethics is not a kind of ethics, and it is not an acceptable replacement for ethics. Population values have an important place *within* ethical reflection; they should not compete with it.

Of course health policy-makers are interested in people's values, and for WHO, in average global health state valuations and variations across populations. Where relevant, these empirical values may be modified in light of UN values relating to basic human rights. For values relating to age weighting and discounting, there are ethical and theoretical concerns that must be taken into account as well as empirical data. In particular, it is not possible to ask people in the future about their health values, though future health states contribute to health gaps in particular.

One notable exception to the argument against a strictly empirical approach to defining key value choices in summary measures relates to the understanding of health state valuations as reflecting different weights on the core domains of health. Unlike values such as time preference, there are no compelling philosophical arguments as to the relative importance of mobility versus cognition in overall assessments of health levels. Thus, we believe it is reasonable to use global average health state valuations in the construction of summary measures and to perform sensitivity analyses using the empirical range of valuations across Member States.

GOODNESS AND FAIRNESS

Part 14 of this volume presents an overview of some of the key debates around fairness concerns in applications of summary measures in health policy. While our view of summary measures as pertaining strictly to levels of health tends to exclude problems of equity and fairness from the actual measures themselves, there are a host of vital philosophical questions relating to the interface between goodness and fairness in societal decision-making. Chapter 14.1 and the commentaries that follow provide a brief introduction to these issues, and we refer readers to a forthcoming volume entitled *Fairness and Goodness* for a much more extensive treatment of the subject (Wikler and Murray, forthcoming).

TIME AND AGE

The incorporation of time preference and age weights in summary measures of population health has been controversial (see part 13). The following section summarizes some of the key arguments and recommendations that emerge from the various debates on these values.

Discounting health

When health gaps are calculated in order to assess the contribution of different diseases, injuries and risk factors to patterns of population health, a major purpose is to inform priorities for policy debate. Decisions on resource allocation require information on many other things as well such as the costs and benefits of specific intervention options, considerations of reducing health inequality or enhancing responsiveness and procedural fairness (see below). Nevertheless, the magnitude of different causes can guide the direction of policy debates about research and development and policy attention itself. As such, the time horizon of the assessment of health gaps should map to the time horizon of the impact of current decision-making. Given this linkage, it is reasonable to focus on the contribution of diseases, injuries and risk factors to health gaps in the near future as opposed to the distant future. The focus on the near future needs to be operationalized by defining a clear time cutoff beyond which health events are not included in the calculation and/or the inclusion of a discount rate in the calculation of health gaps.

In the Global Burden of Disease Study 1990, a 3% discount rate was used in standard DALY calculations (Murray and Lopez 1996). Given the simplicity of a constant time discount rate, and its standard use in discounting of health outcomes in cost effectiveness analyses, it seems prudent to continue to use for incidence-based health gaps measures, the constant 3% time discount rate embodied in the DALY. We recommend, however, given continuing debate about this value, that sensitivity analyses be included in the routine calculation and reporting of health gap measures in order to examine the impact of alternative choices regarding discounting.

For health expectancies, particularly period measures which encapsulate the experience of a representative individual who experiences current period transition rates and/or health state prevalences at each age of their life, time discounting is not appropriate. Likewise, when SMPH are reformulated as averages of individual healthy life expectancies or of individual expected gaps, then there appear to be few reasons to apply time discounting. The most cogent argument for time discounting, the research paradox, can be accommodated by applying generational discounting to future generations, and adding up health or health losses for individuals currently alive without discounting.

Age weights

Standard DALYs in the GBD 1990 included non-uniform age weights in order to reflect the dependency of children and older people on young

adults. This argument, however, revolves around the contributions of young adults to societal well-being, perhaps outside of the realm of population health, strictly defined. Nevertheless, there are other arguments for non-uniform age weighting based on the intuitive notion that lifetimes have a shape, reflecting the way that individuals attach different values to different periods of their own lives. The available evidence does support the proposition that individuals attach different value to different periods of their own lives. In particular, in terms of the comparative judgement of whether population A is healthier than population B, individuals appear to place more weight on the health of individuals at middle ages than at older ages, and possibly at very young ages. While the exact shape of the age weighting function may be disputed, it seems to us useful to continue to compute health gap measures with non-uniform age weights for continuity in comparisons, but to adopt uniform age weights as the standard formulation for health expectancies. As with discounting, we recommend that sensitivity analyses be included in the routine calculation and reporting of health gap measures in order to examine the impact of alternative choices regarding age weights.

MEASUREMENT

The ultimate inputs into summary measures of population health are estimates, by age and sex, of the distribution of the population across different states of health, including full health, on the causes of states of ill-health by age and sex, as well as age-sex-cause-specific mortality rates for each population of interest. This section describes the basic requirements for the calculation of health expectancies and health gaps, and discusses some of the over-riding measurement challenges, in particular relating to the need for cross-population comparability.

MEASUREMENT OF MORTALITY

Reliable estimates of population numbers and mortality rates by age and sex are required for calculating both health expectancies and health gaps for a population. The “gold standard” for population-level data on the numbers of deaths in a population is complete registration of all deaths, by sex and age, and by year of occurrence for the normal resident (*de facto*) population. Each death recorded in a vital registration system should be medically certified by a qualified medical practitioner using the latest version of the International Form of Medical Certificate of Cause of Death. Selection of the cause to be coded as the underlying cause of death should be based on the rules and procedures specified in the 10th Revision of the International Classification of Diseases, Injuries and Causes of Death (WHO 1992), or on subsequent revisions. Where vital registration is incomplete, standard demographic techniques may be used to adjust for under-reporting of deaths (e.g. intercensal analyses or the Brass Growth-Balance method).

Other data collection systems have also been used successfully to compile data on the level of mortality in the absence of good national vital registration systems. Modules on child and adult mortality can be added to household surveys to estimate mortality levels either directly or indirectly. For example, the Demographic and Health Survey programme (DHS) has yielded comparable estimates of infant mortality and child (under 5) mortality for more than 60 countries since 1985. At other ages, mortality, particularly among working age adults, is a much rarer event than child deaths and hence larger samples are generally required. As for child mortality, levels of adult mortality can be estimated either directly from questions about deaths in the household over the past 12–24 months or indirectly from questions about orphanhood and the survival of siblings. Indirect methods for adult mortality are strongly prone to under-estimate mortality levels, frequently by up to 50–60%. Great care is therefore required when interpreting the results of these methods to estimate adult mortality and direct methods, with appropriate interview probes for misreporting, are likely to yield more useful estimates.

Finally, in situations where vital registration covering the entire population is not feasible, sample registration systems, or networks of disease surveillance points have been used to collect data on defined populations. Provided these “catchment” populations are chosen so as to be nationally representative, and provided the routine demographic and epidemiological surveillance is suitably rigorous, sample surveillance schemes can yield very useful data on age-specific death rates and on causes of death. Sample-based schemes have been successfully used for demographic surveillance in India (0.6% sample) and for epidemiological surveillance in China (1% sample). In the case of the Disease Surveillance Points system in China, cause of death is assigned on the basis of medical records for the deceased, where they exist and/or a structured questionnaire about symptoms experienced prior to death (“verbal autopsy”).

MEASUREMENT OF NON-FATAL HEALTH OUTCOMES

Cross-population comparability

The health state measurement field has dealt rather well with issues of validity and reliability of instruments (see chapter 8.1), and this book does not attempt to add to this body of work. Ensuring the cross-population comparability of results adds a third dimension to survey instrument development. The difference between comparability on the one hand and validity and reliability on the other hand can be illustrated using the example of two thermometers, one of which measures temperature in Celsius and the other in Fahrenheit. Both thermometer measures give valid and reliable measurements of temperature. However, 26 degrees on the Celsius thermometer is not comparable to 26 degrees on the Fahrenheit thermometer. Comparability is fundamental to the use of survey results for development of evidence for health policy but has been under-emphasized in instrument development.

Some people may argue that there is no need for comparability in population health measurement. But the basis of science is comparable measurement: comparison creates possibilities of investigating broad determinants at national and cross-national levels. There is a strong demand at national and international levels for comparisons for these reasons, whether consciously articulated or not. Even where there is no interest in cross-national comparison, there is a need for comparison within countries over time, or across different sub-populations delineated by age, sex, education, income or other characteristics. Health measurements, particularly for policy-makers, generally only have meaning in context, and context means comparison. For example, a decrease in mortality of 10% over 10 years for a particular cause in a country, may be “good” or “bad” (e.g. if other comparable countries have achieved a 50% reduction with fewer resources). While it is possible to assess health progress purely with time comparisons within a country, relating this progress to health and other social interventions and trends is extremely difficult in the absence of comparison with other populations, simply because there is only one data point (the set of interventions that actually occurred).

We thus conclude that comparison is extremely important, and fundamental to all uses of SMPH. Population health information is largely irrelevant (i.e. uninterpretable) if we do not have it. This is a new and fertile area that SMPH bring into sharp focus, because mortality is comparable, but most measures of health states are currently not.

The fundamental challenge in seeking cross-population comparable measures is that the most accessible sources of data relating to health are self-reported categorical data. When categorical data are used as the basis for understanding quantities that are determined on a continuous, cardinal scale, the problem of cross-population comparability emerges from differences in the way different individuals use categorical response scales (see chapter 8.3). Response category cut-point shift can make crude comparisons of results across populations nearly meaningless, even when exactly the same questions are used, as illustrated by some of the examples provided in chapter 8.1. Recent analyses of surveys containing both self-reported and measured health status levels have documented systematic differences in the interpretation of self-report survey questions according to age, sex, socioeconomic disadvantage and other individual factors (chapters 8.2 and 8.4).

WHO has developed new approaches to solve the problem of comparability of self-report data, and results from the WHO Multi-Country Survey Study carried out during 2000–2001 provide strong evidence that the methods improve cross-population comparability (Üstün et al. 2001).

Health state valuations

Health state valuations provide the critical link between information on mortality and information on non-fatal health outcomes in summary measures. As reviewed in parts 9 to 11 of this book, the empirical basis for health state valuations has been relatively limited to date, and research has been conducted primarily in the United States and European countries. A key objective for WHO has been to enhance the empirical basis for health state valuations by collecting population-based measurements throughout the world using a set of standardized instruments and protocols based on the best current methods.

During 2000–2001, WHO conducted household and postal surveys on health in more than 60 Member States (Üstün et al. 2001) including a health state valuation component in the 10 large-scale household surveys. Because the available instruments are imperfect, a two-tiered data collection strategy has been used. Many of the techniques developed for the measurement of health state valuations have been designed to be implemented among highly educated respondents, as they rely on abstract and cognitively demanding thought experiments. The only appropriate valuation tool for respondents across a wide range of levels of educational attainment is the visual analogue scale, which has been implemented in diverse cultural settings, including Cambodia, Colombia, India, the Philippines and the United Republic of Tanzania, and has been demonstrated consistently to have higher reliability than other methods (see, for example, chapters 9.3 and 11.3). Based on long-standing results from both psychophysics and psychometrics, however, it is necessary to rescale valuation results obtained through visual analogue scales in order to obtain interval-scaled valuations for construction of summary measures. This rescaling should be empirically-based and therefore requires a second avenue of data collection using deliberative protocols based on multiple states and multiple valuation methods.

While we noted above that it is not necessary to conceptualize health state valuations using the language of choice, preferences do provide a convenient way to elicit information on these valuations. Choice-based elicitation techniques, such as the standard gamble and time trade-off methods, allow inference about levels of health by having respondents weigh changes in these health levels against other quantities, such as mortality risks or length of life. The analytical challenge lies in recovering the underlying assessments about health levels from these responses, which requires a parsing out of the different values behind individuals' answers to these various questions. We have implemented this approach as part of the WHO surveys in 10 countries, using groups with higher levels of educational attainment. The results of the deliberative, multi-method exercises are currently being used to estimate the relationship between responses on a visual analogue scale and the cardinal scale of health state valuations that is required in summary measures (see chapter 9.4).

A number of measurement challenges relating to health state valuations provide a fertile area for continuing research. Examination of variation in valuations across and within countries remains a key priority. To that end, WHO will continue to pursue an ambitious range of data collection activities in representative population samples in diverse settings. Thus far, these efforts have indicated minimal cross-country differences, but we await further results from additional countries. Comparisons between large population-based surveys and small, deliberative convenience samples should be undertaken in order to evaluate the relative measurement characteristics of these alternative modes.

The work summarized in this volume highlights several other important areas for methodological development, including the use of multi-method protocols to estimate the underlying valuations informing responses on standard elicitation techniques, alternative modes of presenting health states as stimuli for valuations, and estimation of the mapping functions that link multi-dimensional health state descriptions and health state valuations. A range of statistical problems must be addressed, including the proper specification of measurement error on health state valuations, the correct model specification for the valuation function and more sophisticated Bayesian modelling strategies for understanding heterogeneity in individual responses to different measurement methods.

Combining health surveys with epidemiological estimates

Health expectancies require information on the prevalence and severity distribution of various health states in the population, by age and sex, which may be obtained, in principle, from suitably large, nationally representative surveys covering the key domains of health. For the causal decomposition of health gaps, it is necessary to map prevalence (and incidence) of health states to disease and injury causes. It is not possible to do this using self-report health survey data, since respondents are often not aware of the causes of their health states, particularly when there are multiple health conditions.

For this reason, it is necessary to collect epidemiological data on specific diseases and injuries in order to ensure internal consistency with mortality data and population level health state prevalence data. Ideally, estimates of the prevalence of disease and injury outcomes (sequelae), by age and sex, would be built up from age-specific incidence, remission and case-fatality rates for each sequela in a given population.

The selection of specific diseases and injuries, the data sources for which might be given greater priority in a broader epidemiological data system, needs to be based on a set of clear criteria. Foremost among these would be diseases or injuries which *a priori* are considered to be major causes of disease burden and/or which are likely to collectively account for, say, two-thirds or more of the disease burden in a population. For each such disease or injury, a consultative process among epidemiologists, health statisticians and other public health personnel concerned with data collection should

be organized to agree upon the ideal and minimal data sources required for calculating summary measures of population health. Sources of such information and their advantages and limitations are discussed in chapters 7 and 10 of Mathers et al. (2001a). Typically, these will include:

1. household surveys of prevalence (e.g. smear positive sputum for tuberculosis, infertility, affective and anxiety disorders, lung function tests, etc.) or of episodes of illness (e.g. diarrhoeal diseases) or seroepidemiological surveys (e.g. for hepatitis B/C);
2. sentinel sites (e.g. antenatal clinics for HIV), incidence of complications of disease (e.g. liver complications, cerebral malaria, amputations and neuropathies);
3. population-based incidence registries (e.g. for cancer, cerebrovascular disease, congenital anomalies, homicide); and
4. hospital discharge data (e.g. for cirrhosis of the liver, fires, road traffic accidents).

Recommendations about the key diseases and injuries for which data are required in order to calculate SMPH are under development by WHO in consultation with Member States. Both the ideal data sources and the minimal data sources are being identified to guide countries to determine priorities for health information systems to facilitate the calculation of these measures.

APPLICATIONS

MEASURING LEVELS OF HEALTH: THE NEED FOR BOTH TYPES OF SMPH

Improving overall levels of population health is a clear priority for WHO Member States. Healthy life expectancy is a readily understandable positive measure that is appropriately sensitive to survival rates and to the prevalence and severity distribution of health states among the population. As healthy life expectancy is based on a severity distribution of prevalence of health states, it is more comparable across populations than other measures based on dichotomous prevalence measures, since distributional data imply that the measure is much less sensitive to the threshold chosen for defining ill-health.

Neither life expectancy nor healthy life expectancy provide information on the leading causes of death or non-fatal health status in populations. It is not possible to disaggregate these measures in an additive fashion by cause (see chapter 4.1). Health gaps are additive across causes and across population groups and hence provide a more appropriate and useable SMPH for reporting on causes of loss of health for Member States (see chapters 5.1 and 6.1).

Summary measures are only one component of a coherent and integrated statistical framework. SMPH should be viewed as the apex of a hierarchy of related measures, rather than a piecemeal set of unconnected measures. The macro measures at the apex of the system, such as health expectancies and health gaps, provide a broad population-based overview of levels and causes.

SMPH AND SUMMARY MEASURES OF HEALTH INEQUALITY

Information about average levels of health in a population is important for health policy formulation, as is knowledge about inequalities in health within a population. Should there be one summary measure of population health that reflects both average levels of population health and the distribution of health within the population, or should there be separate summary measures of average population health and health inequality?

Some values such as concern for the distribution of health outcomes may figure prominently in public decision-making, but it is still debatable whether such fairness concerns should be integrated directly into the summary measure of population health or rather measured independently. The advantages of including fairness concerns in a summary measure is that it places these issues firmly in the health agenda. The disadvantage is that including fairness considerations directly in a summary measure of population health can complicate the summary measure profoundly and does not allow for different trade-offs between goodness and fairness.

Perhaps more importantly, it is clear that there may be fairness concerns that are central to the choice of interventions but are not as relevant to the comparative use of summary health measures. Nord (chapter 14.4) has drawn attention to a preference for giving health resources to the sick to some extent irrespective of their capacity to benefit, a form of distributional concern. One cannot argue that such a priority to the sick would be relevant to measuring population health even if it is critical to the debate on resource allocation across interventions and beneficiaries. Keeping fairness and goodness considerations distinct in the construction of summary measures of population health allows us to keep track of these different uses and needs. In general, a much sharper distinction needs to be made in the debate on the construction of summary measures of population health and the ethical dimensions of intervention choice. Some values such as discounting or age-weighting may be considered types of fairness considerations (discounting is related to intergenerational fairness and age-weighting to fairness across age groups). Or they may be seen as components of goodness (in this case amount of health). If the latter, there is a much stronger case that they be incorporated into a summary measure of population health.

Precedents for separate summary measures exist from other sectors. For example, in the case of income measurement, the common practice is to report average income per capita for a population and separately to report a measure of income inequality such as the Gini coefficient. There is,

nevertheless, a rich literature on composite measures of average income and its distribution (e.g. Kolm-Atkinson measures, etc.). For health, it is simpler, closer to the tradition of health statistics and more easily communicated to the general public to have separate summary measures of average population health and health inequality. Consequently, we believe that WHO should develop and report on separate summary measures for level and distribution of health in WHO Member States.

The desirability of two distinct measures for summarizing the average levels of population health and health inequalities has implications for methods used to measure the valuations for time spent in health states less than full health. These methods should not incorporate distributional concerns. Rather, these distributional concerns should be directly captured in the summary measure of health inequality.

SMPH AND RESOURCE ALLOCATION

What is the use of summary measures of population health decomposed into the contributions of diseases, injuries or risk factors to prioritising investments in different interventions? Williams (1999) and Mooney (chapter 12.3 in this book) have claimed that a purpose of SMPH is to set health funding priorities and argued that this is inappropriate. Neither we, nor our colleagues, have ever claimed that resources should be directed toward health problems solely on the basis of their relative contributions to premature mortality and non-fatal health outcomes. Williams and Mooney are constructing a straw man by implying that we advocate using SMPH alone to select funding priorities.

While nearly everyone agrees that one very important input to prioritising resources for interventions is the cost-effectiveness of interventions, nevertheless, debates on priorities for health action can be informed by summary measures of population health. To estimate the benefits of an intervention at the population level and to monitor the impact of specific interventions when implemented, a valid assessment of the epidemiology of the disease, injury or risk factor addressed by the intervention is required. There are very few instances where marginal change in health at the population level can feasibly be measured in any way other than in terms of the difference between two assessments of level of health using SMPH. Additionally, where there are non-monetary fixed assets limiting the feasible combinations of interventions that can be delivered, such as the attention of senior Ministry of Health decision-makers, then these assets should be devoted not just to the most cost-effective interventions but to those cost-effective interventions with the potential to effect substantial improvements in population health status.

Information on the contributions of diseases, injuries and risk factors to summary measures of population health is also necessary for the health intelligence function of governments. If a disease, injury or risk factor is not yet recognized as a major problem, there will be no attempt to formulate intervention strategies or even to assess the benefits and costs

of alternative intervention strategies. For example, one consequence of the GBD 1990 study has been to focus policy attention on intervention strategies for depression in low and middle-income countries, resulting in efforts to analyse the cost-effectiveness of interventions for depression. If decomposition of summary measures into the contributions of diseases, injuries and risk factors is necessary, this has significant implications for the design of summary measures, favouring health gaps over health expectancies (see Murray et al. chapter 5.1 in this volume).

The argument made by Williams and Mooney—that only incremental changes through intervention, and not the level of population health, matter—appears strange when extended by analogy to national income and product accounts. As Jamison (1996) writes on this analogy, “The most natural comparison [to the GBD] is to the development of National Income and Product Accounts (NIPAs).” (p. xix) Original pioneering efforts on NIPAs were followed by the codification of international standards in the System of National Accounts (SNA 1993). Despite codification, debate has continued unabated on the conceptual and empirical basis for national accounts—should environmental degradation be included in capital depreciation, should household production be included, etc. National accounts measure the level of economic activity in a country. Williams and Mooney must surely argue that this is unnecessary; they must argue that resources wasted on measuring national accounts could be better spent on calculating the incremental gains in national income that could be achieved through various policy options or interventions. Today, the myriad uses of national accounts have so enriched the field of macroeconomics that little energy is spent on questioning their utility. Summary measures of population health are, for the health sector, a natural analogue to national income and product accounts.

It is essential to distinguish clearly efforts to quantify the health gains from interventions for cost-effectiveness analyses from efforts to apply summary measures of population health in cross-national comparisons or other uses. Imagining that the most desirable summary measure of population health has been identified, one could argue logically that the benefits of a health intervention should be measured as the expected difference in this summary measure for a population with and without the intervention (Murray and Lopez 2000). At minimum, there should in general be consistency between the approach used to develop summary measures and that used to estimate the benefits of interventions, without necessarily formally defining the benefits as the change in the summary measure. Thus, for example, health state valuations might be the same for both uses, but the benefits of interventions could be measured directly as the increase in healthy years of life lived.

CONCLUDING REMARKS

In this concluding chapter, we have attempted to bring together the various issues and viewpoints raised by the contributors to this volume in a way that will be helpful for policy and further scientific debate. These issues range from the conceptual underpinnings of summary measures of population health, to the technical and practical details of their construction and use, and to the various ethical considerations that such measures unavoidable entail. We have not tried to be prescriptive on all issues: indeed from the diversity of views and opinions expressed in this book, it would be premature to do so. Rather, in this chapter we have tried to draw together these opinions into coherent scientific argument around each of the key themes we have identified in the construction and use of summary measures. This should greatly facilitate the application of this body of work in public policy and better focus the agenda for future research. Where there is consensus, or virtual consensus, we have tried to be very explicit about the implications for summary measures of population health, and where there is not, we have tried to lay out the issues in a constructive and objective fashion. Transparency will increase confidence that a broad spectrum of legitimate scientific opinion has been canvassed and thus enhance the policy relevance of its content.

The Global Burden of Disease project, and subsequent work on burden of disease analysis and the use of healthy life expectancy to measure population health attainment in the analysis of health system performance of WHO Member States (WHO 2000; WHO 2001b) has stimulated broader interest in the design, calculation and use of summary measures of population health. The body of work on the burden of disease and comparative risk assessment has also demonstrated the feasibility of estimating health gaps and their key role in quantifying the contribution of diseases, injuries and risk factors for population health. Notwithstanding these contributions, there is clearly further scope for improvement in all aspects of the development and use of SMPH. The World Health Organization through its commitment to routinely assessing the global burden of disease, to regularly assessing the performance of health systems of Member States, and to leading an international dialogue on the development of summary measures of population health, has laid the scientific and institutional foundations to foster and advance this agenda. Meanwhile, we trust that the substantive debate and presentation of ideas collected in this volume will serve to promote greater understanding of the scope and potential of summary measures of population health to inform public health development and health policy.

REFERENCES

- Boorse C (1977) Health as a theoretical concept. *Philosophy of Science*, 44:542–573.
- Jamison DT (1996) Foreword to the global burden of disease and injury series. In: *The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries and risk factors in 1990 and projected to 2020. Vol 1*, Murray CJL, Lopez AD, eds., Harvard School of Public Health on behalf of WHO, Cambridge, MA.
- Mathers CD, Vos T, Lopez AD, Salomon JA, Ezzati M, eds. (2001a) *National burden of disease studies: a practical guide. 2.0.*, World Health Organization/Global Programmememe on Evidence for Health Policy, Geneva. www.who.int/evidence/nbd
- Mathers CD, Sadana R, Salomon J, Murray CJL, Lopez A (2001b) Healthy life expectancy in 191 countries, 1999. *Lancet*, 357(9269):1685–1691.
- Murray CJL, Lopez AD, eds. (1996) *The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries and risk factors in 1990 and projected to 2020. Vol 1.*, Harvard School of Public Health on behalf of WHO, Cambridge, MA.
- Murray CJL, Lopez A (2000) Progress and directions in refining the global burden of disease approach: a response to Williams. *Health Economics*, 9(1):69–82.
- Murray CJL, Lopez AD (1999) On the comparable quantification of health risks: lessons from the global burden of disease study. *Epidemiology*, 10(5):594–605.
- Robine JM, Jagger C, Romieu I (2001) Disability-free life expectancies in the European Union countries. *Genus*, 57(2):89–101.
- Sen A (1992) *Inequality re-examined*. Russel Sage; Harvard University Press, New York; Cambridge, MA.
- SNA (1993) *System of national accounts 1993*. European Communities, International Monetary Fund, Organisation of Economic Cooperation and Development, United Nations, World Bank. <http://esa.un.org/unsd/sna1993/>
- Üstün TB, Chatterji S, Villaneuva M, Bendib L, Sadana R, Valentine N, Ortiz J, et al. (2001) *WHO multi-country household survey study on health and responsiveness, 2000–2001*. GPE discussion paper no 37. World Health Organization/Global Programmememe on Evidence for Health Policy, Geneva.
- WHO (1948) *Constitution*. World Health Organization, Geneva.
- WHO (1992) *International statistical classification of diseases and related health problems (ICD 10)*. 10th edn, World Health Organization, Geneva.
- WHO (2000) *The World Health Report 2000. Health systems: improving performance*. World Health Organization, Geneva.
- WHO (2001a) *International classification of functioning, disability and health (ICF)*. World Health Organization, Geneva.
- WHO (2001b) *The World Health Report 2001. Mental health: new understanding, new hope*. World Health Organization, Geneva.

Wikler D, Murray CJL, eds. (Forthcoming) *Fairness and goodness: ethical issues in health resource allocation*. World Health Organization, Geneva.

Williams A (1999) Calculating the global burden of disease: time for a strategic reappraisal? *Health Economics*, 8(1):1–8.