

# A critical examination of summary measures of population health

Christopher J.L. Murray,<sup>1</sup> Joshua A. Salomon,<sup>2</sup> & Colin Mathers<sup>3</sup>

In the past decade, interest has been rising in the development, calculation and use of summary measures of population health, which combine information on mortality and non-fatal health outcomes. This paper reviews the issues and challenges in the design and application of summary measures and presents a framework for evaluating different alternatives. Summary measures have a variety of uses, including comparisons of health in different populations and assessments of the relative contributions of different diseases, injuries and risk factors to the total disease burden in a population. Summary measures may be divided into two broad families: health expectancies and health gaps. Within each family, there are many different possible measures, but they share a number of inputs, including information on mortality, non-fatal health outcomes, and health state valuations. Other critical points include calculation methods and a range of conceptual and methodological issues regarding the definition, measurement and valuation of health states. This paper considers a set of basic criteria and desirable properties that may lead to rejection of certain summary measures and the development of new ones. Despite the extensive developmental agenda that remains, applications of summary measures cannot await the final resolution of all methodological issues, so they should focus on those measures that satisfy as many basic criteria and desirable properties as possible.

**Keywords:** health status indicators; cost of illness; epidemiologic studies; evaluation studies.

*Voir page 991 le résumé en français. En la página 992 figura un resumen en español.*

## Introduction

Summary measures of population health combine information on mortality and non-fatal health outcomes to represent the health of a particular population as a single number (1). Efforts to develop such measures have a long history (2–10). In the past decade there has been a markedly increased interest in the development, calculation and use of summary measures. The volume of work from members of the *Réseau de l'Espérance de Vie en Santé* (REVES) offers one indication of the activity in this field. (11, 12). Applications of measures such as active life expectancy (ALE) (13) have been numerous, especially in the USA. Calculations of related summary measures such as disability-free life expectancy (DFLE) have also appeared frequently in recent years (14–19). Another type of summary measure, disability-adjusted life years (DALYs), has been used in the Global Burden of Disease Study (20–26) and in a number of

national burden of disease studies (27–36). WHO is committed to routine measurement and reporting of the global and national burdens of disease (37, 38). The United States Institute of Medicine (IOM) convened a panel on summary measures and published a report that included recommendations for enhancing public discussion of the associated ethical assumptions and value judgements, establishing standards and investing in education and training to promote the use of such measures (1).

We review below the range of options for summarizing population health, and the main challenges and debates underlying them. Because there are many options, we propose criteria that can be used to evaluate different summary measures of population health. The intended use of a measure may have important implications for its design, and we therefore outline the major uses of summary measures. A brief discussion of the information requirements for all summary measures is followed by a typology of summary measures in terms of health expectancies and health gaps. We outline key issues of importance for all summary measures. A number of criteria and other properties are proposed which can be used to evaluate different summary measures. We discuss some of the broad implications of this framework for choosing summary measures and consider the prospects for future progress.

<sup>1</sup> Director, Global Programme on Evidence for Health Policy, World Health Organization, 1211 Geneva 27, Switzerland. Correspondence should be addressed to this author.

<sup>2</sup> Health Policy Analyst, Global Programme on Evidence for Health Policy, WHO.

<sup>3</sup> Scientist, Global Programme on Evidence for Health Policy, WHO.

This paper should be understood in the context of work under way in WHO on the development of an analytical framework for measuring health system performance (39). We consider one critical element of this framework, namely the need for measures of population health that capture the average levels of fatal and non-fatal health outcomes in a population. WHO is also developing measures that summarize health inequalities in populations (40, 41). Assessments of health systems thus depend on both summary measures of the average level of population health and measures of the distribution of health among individuals.

## Uses of summary measures

The design of a summary measure may depend on its intended use. Some potential applications are indicated below.

1. *Comparing the health of one population with that of another.* Such comparisons are an essential input into evaluations of the performance of different health systems, along with information on health inequalities, responsiveness and fairness in financing (39). Comparisons may allow decision-makers to focus their attention on health systems with the worst achievement for a given level of resources. Comparative judgements also provide the dependent variable in analyses of the independent variables that contribute to health differences between populations.

2. *Monitoring changes in the health of a given population.* Monitoring changes in health status over time is essential in the evaluation of health system performance and progress towards stated goals for a given society.

3. *Identifying and quantifying overall health inequalities within populations.*

4. *Providing appropriate and balanced attention to the effects of non-fatal health outcomes on overall population health.* In the absence of summary measures, conditions that cause decrements in function but not mortality tend to be neglected in favour of conditions that primarily cause mortality.

5. *Informing debates on priorities for health service delivery and planning.* A summary measure can be combined with information on the contributions of different causes of disease and injury or risk factors to the total. Such information should be a critical input to debates on the identification of a short list of national health priorities that will receive the attention of senior managers in public health agencies and of government leaders.

6. *Informing debates on priorities for research and development.* The relative contributions of different diseases, injuries and risk factors to the total summary measure also represent a major input to the debate on priorities for investment in research and development (42).

7. *Improving curricula for professional training in public health.*

8. *Analysing the benefits of health interventions for use in cost-effectiveness analyses.* The change in some summary measure of population health offers a natural unit for quantifying intervention benefits in these analyses.

Consideration of the intended use of summary measures of population health, whether for simple comparative purposes or more tailored policy debates, can be expected to figure centrally in the development of criteria for evaluating alternative summary measures. Nevertheless, in examining the properties of various summary measures it is important to recognize that all applications, even simple comparative ones, can influence the policy process (1). Because of their potential influence on international and national decisions relating to the allocation of resources, summary measures should be considered to be normative. As stated by the IOM panel, all measures of population health involve choices and value judgements in both their construction and their application. Great care must be taken in the construction of summary measures precisely because they may have far-reaching effects. Normative aspects of the design of summary measures continue to be the subject of extensive debate (43–48).

## Information requirements for summary measures

It is important to distinguish clearly between the nature and quality of various inputs to summary measures and the properties of the measures themselves. Information on age-specific mortality and the epidemiology of non-fatal health outcomes provides a basic input to any type of summary measure. Another critical input is information on the values attached to various health states relative to ideal health or death. Instruments for the measurement of health states, such as SF-36 (49), can be used to describe health states in terms of performance in various domains of health. This information could be combined with valuations of health states in order to calculate the non-fatal health component of many different summary measures, but SF-36 and other instruments for the measurement of health states are not in themselves summary measures of population health. The occasional confusion surrounding the distinction between data inputs to summary measures and the summary measures themselves may be exacerbated when summary measures are linked by definition to particular health status instruments, as in years of healthy life (YHL) (50).

## A typology of summary measures

A wide array of summary measures has been proposed. On the basis of a simple survivorship curve they can be divided broadly into two families: health expectancies and health gaps. The green curve in Fig. 1 is an example of a survivorship curve for a

hypothetical population. This curve indicates, for each age along the x-axis, the proportion of an initial birth cohort that will remain alive at that age.

Area  $A + B$  under the bold survivorship curve represents life expectancy at birth. Health expectancies are measures of this area which take into account some lower weights for years lived in health states worse than full health, represented as area  $B$  in the diagram. More formally:

$$\text{health expectancy} = A + f(B),$$

where  $f(\cdot)$  is a function assigning weights to health states less than ideal health, using a scale on which full health has a weight of 1.

A wide range of health expectancies has been proposed since the original notion was developed (2), including ALE (13), DFLE (11, 12), disability-adjusted life expectancy (DALE) (34), YHL (50), quality-adjusted life expectancy (QALE) (7, 51), dementia-free life expectancy (52) and health capital (53, 54).

In contrast to a health expectancy, a health gap quantifies the *difference* between the actual health of a population and some stated norm or goal for population health. The health goal implied in Fig. 1 is for everyone in the population to live in ideal health until the age indicated by the vertical line enclosing area  $C$  at the right. The health gap shown in Fig. 1 can be interpreted as either the life table health gap, i.e. the health gap of a hypothetical birth cohort exposed to a set of currently measured mortality and non-fatal health outcome transition rates, or the absolute health gap of a stable population with zero growth.

Since Dempsey (55), there has been extensive development of various measures of years of life lost attributable to premature mortality (e.g. 44, 56). Measures of years of life lost are all measures of a mortality gap, the area between the survivorship function and an implied population norm for survivorship, represented as area  $C$  in Fig. 1. Health gaps extend the notion of mortality gaps to account for time lived in health states worse than ideal health. The health gap, therefore, is a function of areas  $C$  and  $B$ , or, more formally:

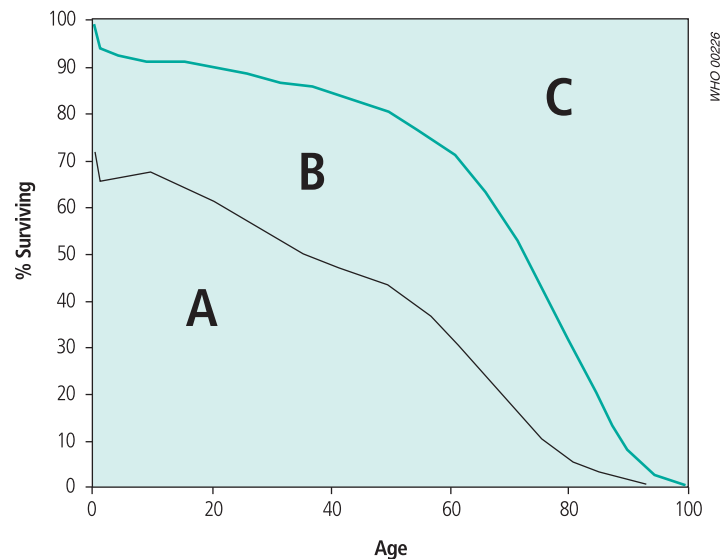
$$\text{health gap} = C + g(B),$$

where  $g(\cdot)$  is a function assigning weights to health states less than full health, using a scale on which a weight of 1 implies that time lived in a particular health state is equivalent to time lost because of premature mortality. Various health gaps have been proposed and measured (9, 38, 44, 57), and many others could be derived.

## Key issues in the design of summary measures

There are at least four sets of issues that cut across all summary measures of population health: technical issues of calculation, the definition and measurement of health states, the valuation of health states, and the inclusion of other social values.

Fig. 1. The survivorship curve



## Calculation methods

### *Absolute and covariate-independent forms of summary measure.*

While the survivorship function in Fig. 1 provides a convenient heuristic illustration of the difference between health expectancies and health gaps, it is important to recognize that summary measures may take either an absolute or an age-independent form. For example, the number of deaths in a population is an absolute measure, while a period life table does not depend on the age distribution in a population. By their construction, all health expectancies are measures that do not depend on the particular age structure of a population. Health gaps, on the other hand, are usually expressed in absolute terms and as such are dependent on age structure. For example, a health gap may be expressed as the total number of healthy years of life that have been lost in a population; this varies with the age distribution of the population. A life table health gap as illustrated in Fig. 1 can be calculated easily and is independent of the age structure of the population. It is also possible to conceive of health expectancies that depend on age structure, e.g. total healthy years lived in a population, but these measures have not been developed thus far.

Most discussions of health statistics have focused on the development of measures that are independent of age structure. Clearly, age is only one of innumerable covariates of health outcomes, and we might therefore imagine developing sex-independent, race-independent or income-independent forms of summary measure just as we have forms that are independent of age structure. There are arguments, however, for paying special attention to the latter:

- age is one of the most powerful determinants of health outcomes, so that comparisons based on measures that are not independent of age structure may be dominated, in some cases, by variation in this variable;

- age cannot be changed by an intervention;
- age is unique in that all individuals belong successively to every age until they die.

Although it is possible to imagine applying at least some of these arguments to other factors, such as sex and race, there will probably always be a particular interest in summary measures that are independent of age structure. The design of a summary measure and the range of covariate-independent forms of the measure that might be developed depend ultimately on its intended use.

**Calculation of health expectancies.** As with standard life tables (58), health expectancies can be calculated for a period or for a cohort. The first method, which is more common, calculates the health expectancy for a hypothetical birth cohort exposed to currently observed event rates (e.g. rates of mortality, incidence and remission) over the course of its lifetime. We are not aware of any calculations of cohort health expectancies for real populations, although longitudinal studies may provide such opportunities (59). Deeg et al. (60) projected disability transition rates based on longitudinal data for the Netherlands but did not convert them into cohort health expectancies. Barendregt & Bonneux (61) calculated changes in cohort disease-free life expectancy attributable to hypothetical interventions in a simulation model.

Health expectancies may also be distinguished by the use of incidence or prevalence information on non-fatal health outcomes. The pioneering efforts of Sullivan (5) and others to estimate health expectancies involved applications of the prevalence rate life table borrowed from working life, marriage and education life tables (e.g. 62). Katz et al. (13) proposed that calculations of active life expectancy should be based on double decrement life tables, in which individuals can move into two absorbing states: limited function and death. More recently, multistate life tables have been estimated for health expectancies (63–65). Robine et al. (66) and Barendregt et al. (67) argued that the multistate method was required logically so that health expectancy would be based only on currently measured mortality, incidence and remission, and not on prevalence. Robine et al. (66) argued that prevalence was not a period measure; it was a stock variable rather than a flow. In real populations there may not be much difference between health expectancies calculated using the prevalence, double decrement or multistate approaches (68).

**Calculation of health gaps.** The most fundamental consideration in the calculation of health gaps is the choice of a target or norm for population health. Health gaps measure the difference between current conditions and a selected target. The explicit or implicit target is a critical characteristic of any health gap. Despite the obvious importance of choosing the health target, in some cases the population target is neither stated nor easily calculated. This was true of one of the first health gaps proposed by the Ghana Health Assessment Project Team (9). The original formulation

of many mortality gaps was constructed in terms of the loss to each individual. The aggregate population implications of the loss due to premature mortality for each individual have been poorly appreciated. For example, Murray (44) has shown that for many mortality gap measures and health gaps the implied target may change as the mortality level changes, making direct comparisons between communities impossible.

As with health expectancies, health gaps may be calculated in various ways that reflect differences between period and cohort perspectives and incidence or prevalence perspectives. For example, DALYs in the Global Burden of Disease 1990 Study have been calculated in two ways: using the incidences of mortality and non-fatal health outcomes, and using the incidence of mortality and the prevalence of non-fatal health outcomes. Still other combinations are possible. Healthy life years (HeaLYs) for a given time period are calculated on the basis of the incidence of pathological processes and the future non-fatal health outcomes and mortality from those processes (57). A pure prevalence health gap could be constructed based on the prevalences of non-fatal health outcomes and of individuals who have died in the past and would have lost years of life in the present time period.

### Definition and measurement of health states

There is an important source of variation across summary measures in the definition and measurement of health states worse than perfect health. How should the different health states comprising area B in Fig. 1 be described? Critical issues have to be considered even before the psychometric properties of different measurement instruments for health are analysed. Among them are the domains of health which are measured, the difference between performance and capacity in a domain, and the determinants of discrepancies between self-reported and observed performance or capacity in a domain.

Many domains of health have been proposed, ranging from each of the senses, to pain, mobility and cognition, and finally to complex functions related to health, such as social interaction or usual activities (69). The International Classification of Impairments, Disabilities and Handicaps (70) attempts to classify this broad range of health domains into body functions, activities and participation. Measurement instruments also vary as to whether they focus on individuals' capacity to perform in a domain, as with the Health Utilities Index, or their actual performance. As many commonly used instruments depend largely on self-reporting, the individual social, economic and cultural factors that influence expectations for performance or capacity on each domain can lead to substantial deviations between self-reported values and observed values (71).

Many health expectancies are linked to a particular instrument for the measurement of health status. ALE is linked to variants of the activities of daily living. The YHL measure is linked to two

questions collected in the United States National Health Interview Survey, concerning activity limitations and perceived general health (50, 72). QALE (51) is linked to a question on activity restriction in the Canada Health Survey. In other cases, such as dementia-free life expectancy, health expectancy is linked to a particular diagnosis or a single domain of health. DFLE is often calculated from data on long-term disability and includes the duration of a condition in its definition of disability. Data on self-assessed general health have been used in the calculation of health capital (53), although the measure is not, by definition, linked strictly to this instrument. Clearly, wherever a health expectancy is defined with reference to a particular instrument, the summary measure depends critically on the reliability and validity of the instrument. All of the particular instruments mentioned here represent very limited conceptions of health, emphasizing a restricted set of physical domains. This contrasts with other health status instruments more widely applied in current practice, such as EuroQol (73) or SF-36 (49), which capture multiple dimensions of health. Direct linkages are not required in the construction of summary measures, so they may complicate evaluation of the properties of summary measures unnecessarily.

### Valuation of health states

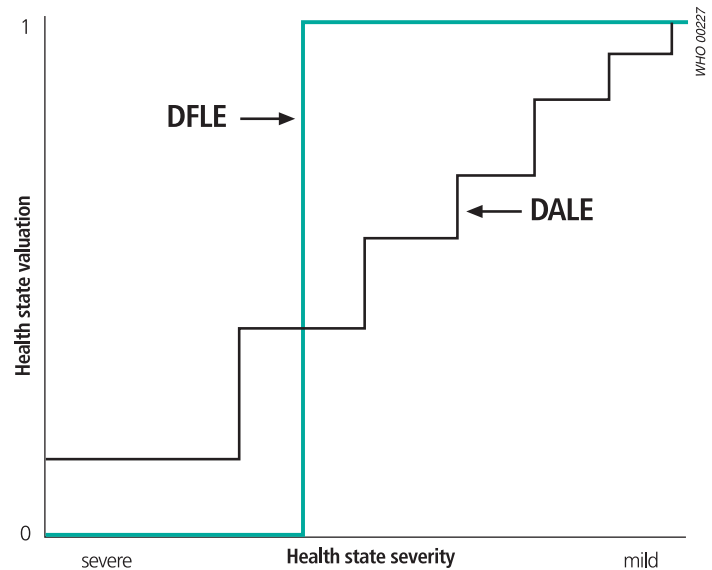
Once the health states represented in area B of Fig. 1 have been described in various domains the next step in calculating either health expectancies or health gaps is to determine the value of time spent in each state relative to full health and death. This allows these non-fatal health outcomes to be combined with information on mortality.

Many health expectancies, such as ALE and DFLE, apply dichotomous valuations (Fig. 2). Up to an arbitrary threshold the valuation is zero (i.e. equivalent to the valuation of death); beyond this threshold the valuation is one (i.e. equivalent to full health). Dichotomous valuations make the measure extremely sensitive to variation in the arbitrary threshold definition, which creates significant obstacles to cross-national comparisons and assessments of changes over time. Other health expectancies and health gaps such as YHL, DALE, health capital and DALYs use polychotomous or, in principle, continuous valuations (Fig. 2).

For those summary measures that do not use arbitrary dichotomous schemes the valuation approach can be distinguished further on the basis of:

- the persons whose values are used, e.g. individuals in health states, relatives of these individuals, the general public or health care providers;
- the type of valuation question that is used, e.g. the standard gamble, time trade-off, person trade-off, or visual analogue;
- the manner of presenting health states for the elicitation of valuations, i.e. with what type of description and what level of detail, including some selection of domains;

Fig. 2. **Valuations of time spent in health states worse than perfect health.** The green line indicates dichotomous valuations, as in disability-free life expectancy, while the black line indicates categorical valuations, as in disability-adjusted life expectancy.



- the range of health states, from mild to severe, valued at the same time;
- the combination of valuation questions, and, more generally, the type of deliberative process undertaken, if any.

The relative merits of each of these choices continue to be debated extensively in the health economics literature.

### Other values

Values other than health state valuations also may be incorporated explicitly into summary measures. For example, health capital (53, 54) includes individuals' discount rates for future health. In addition to discounting, some variants of DALYs (20, 34) have included age weights, which allow years lived at different ages to take on different values. Equity weights have also been proposed (48) in order to allow years lived by one group or another to take on different values. Incorporating other values into the design of a summary measure usually requires strong assumptions about the separability of health across both persons and time. Such separability has been challenged in the literature on quantifying the benefits of health interventions, as in the debate on healthy year equivalents and quality-adjusted life years (74–76).

### Criteria for evaluating summary measures

Given the extensive interest in summary measures and the range of health expectancies and health gaps, one way to proceed is to propose a minimal set of

desirable properties that summary measures should have and to evaluate available summary measures against these criteria. The minimalist set of desirable properties for summary measures is likely to vary with the intended use. Thus a summary measure most appropriate for comparisons of population health over time may not be the most appropriate or even acceptable for reporting on the contributions of diseases, injuries and risk factors to ill health in a population. We consider the choice of an appropriate summary measure for comparative purposes, and then take up the question of choosing a summary measure that can be decomposed into the contributions of different diseases, injuries and risk factors. The purpose is to begin to define an explicit framework for making these choices.

There is a common-sense notion of population health according to which, for some examples, everybody could agree that one population is healthier than another, or that the health of a particular population is becoming worse or better. For instance, if two populations are identical in every way except that infant mortality is higher in one, it is to be expected that everybody would agree that the population with the lower infant mortality is healthier. On the basis of this type of common-sense notion we can develop some very simple criteria for evaluating summary measures of health. However, even simple criteria lead to some rather thought-provoking conclusions.

Much of the discussion on the design of summary measures has been linked closely to the goal of maximizing gain in a summary measure in the face of a budget constraint. Inevitably, this has led to methods for constructing summary measures that emphasize the myriad value choices involved in the allocation of scarce resources, for example the use of the person trade-off technique for measuring health state valuations (44, 77). Many authors have rightly focused on a range of values relevant to the allocation of scarce resources that may enhance individuals' health (78). However, many of these considerations bring us far from the common-sense statement that one population is healthier than another. At least for the purposes of comparative statements on health it may be necessary to distance the development of summary measures from the complex values that have to be considered in the allocation of scarce resources. In other words, we can quite reasonably choose to measure population health in one way and conclude that scarce resources should not be allocated strictly to maximize population health as so measured. Indeed, implicit in the WHO framework for measuring health system performance (39) is the notion that resources should at least be allocated to maximize some socially desired mix of:

- average levels of population health;
- reductions in health inequalities;
- responsiveness of the health system to the legitimate expectations of the public regarding the non-health dimensions of its interaction with the system;
- fairness of health system financing.

Tentatively, we believe that we can construct summary measures for comparative purposes based on an application of Harsanyi's principle of choice from behind a veil of ignorance (79). In this construct, an individual behind a veil of ignorance does not know who he or she is in a population.<sup>a</sup> We propose that the relation "is healthier than" can be defined in such a way that population A is healthier than population B if, and only if, an individual behind a veil of ignorance would prefer to be one of the existing individuals in population A rather than an existing individual in population B, holding all non-health characteristics of the two populations to be the same.<sup>b</sup> We emphasize that the principle of choice behind the veil of ignorance does not mean choosing to join one of the populations as an *additional* member. A person must choose between two populations knowing that he or she would be one of the *current* members of either population, but not knowing at the moment of choice *which particular* member he or she would be in either population. Implementing the veil of ignorance approach to selecting the criteria for a summary measure of population health would have many far-reaching implications. On the basis of the veil of ignorance argument and consonant with common-sense notions of population health we argue that there are, minimally, five criteria that a summary measure should fulfil. These criteria are presented below with examples of comparisons between two populations at an instant in time.

**Criterion 1.** If age-specific mortality is lower at any age, *everything else being equal*, then a summary measure should be better (i.e. a health gap should be lower and a health expectancy should be higher). Strictly speaking, this criterion refers to mortality rates among individuals in health states that are preferred to death. This criterion could be weakened to say that if age-specific mortality is lower at any age, everything else being equal, then a summary measure should be the same or better. The weaker version would allow for deaths beyond some critical age to

<sup>a</sup> Harsanyi's veil of ignorance has been described as a thin veil, in contrast to Rawls's thick veil. In Rawls's formulation (80), the veil excludes much more information; most importantly for this discussion, it excludes the particular circumstances of society such as the epidemiological information upon which our criteria are based.

<sup>b</sup> Formally, the assumption that the relation "is healthier than" does not depend on the particular levels at which non-health characteristics are fixed requires separability of the health-related characteristics of a population from characteristics that are not health-related. This assumption has been questioned (81) and several compelling examples of its violation have been presented. Nevertheless, there are reasons to believe that health is largely separable from other components of well-being. In nearly all languages and cultures there is a recognized word for health and a distinct concept of health. Common sayings to the effect that health is more important than wealth serve as a testament to the basic separability of health and non-health well-being.

In addition to the separability of health and non-health well-being, we require that individuals behind the veil of ignorance make their choice assuming that health is separable across individuals. This is required so that the statement "is healthier than" strictly reflects differences in the average level of health between populations and not the distribution of health within populations. We intend to capture distributional issues in a separate measure of inequality of health across individuals.

leave a summary measure unchanged. Measures such as potential years of life lost would then fulfil the weak form of the criterion.

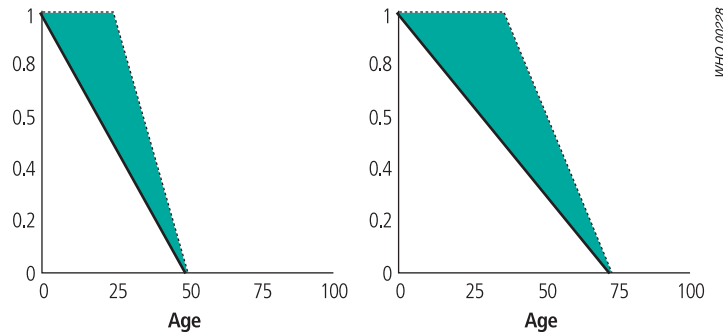
By inspection, all health expectancies fulfil the strong form of this criterion but some health gaps do not. For health gaps, satisfaction of this criterion depends critically on the selection of the normative goal for population survivorship. For example, it can be demonstrated that the use of local life expectancy at each age to define the gap associated with a death at that age, as proposed by several authors (9, 45), leads to a violation of this criterion. For the purposes of illustration, let us imagine two hypothetical populations with linear survivorship functions, as represented by the bold diagonals in Fig. 3. For the population represented in the first diagram, life expectancy at birth (the area under the survivorship function) is 25 years, while the second population has a life expectancy at birth of 37.5 years. Based on the survivorship function,  $s(x)$ , we can compute, for each population, the life expectancy at each age,  $e(x)$ , namely, the area under the survivorship function to the right of age  $x$  divided by  $s(x)$ . The implied population norm,  $G(\cdot)$ , is defined so that  $G(x + e(x)) = s(x)$ . In Fig. 3,  $G(\cdot)$  is shown as the diagonal line to the right of the survivorship curve. In the population with a life expectancy of 37.5, which has a lower mortality rate at every age than the population with a life expectancy of 25, the health gap shown as the shaded area has actually increased.

**Criterion 2.** If age-specific prevalence is higher for some health state worse than ideal health, *everything else being equal*, a summary measure should be worse. Let us imagine two populations, A and B, with identical mortality, incidence and remission for all non-fatal health states but with a higher prevalence of paraplegia in population A. Behind a veil of ignorance an individual would be expected to prefer to be a member of population B. Likewise, the common-sense notion of population health leads us to conclude that B is healthier than A. Health expectancies and health gaps calculated only on the basis of incidence and remission rates for non-fatal health states do not fulfil this criterion, whereas prevalence-based health expectancies and health gaps do fulfil it.

**Criterion 3.** If age-specific incidence of some health state worse than ideal health is higher, *everything else being equal*, a summary measure should be worse. Let us imagine two populations, A and B, with identical mortality, prevalence and remission, but with a higher incidence of blindness in A than in B. We must conclude that B is healthier than A. Incidence-based health expectancies and health gaps would fulfil this criterion.

Taking criteria 2 and 3 together, we are led to the conclusion that no existing summary measure fulfils both of them. According to conventional wisdom in health statistics, incidence-based measures are better for monitoring current trends and are more logically consistent for summary measures because mortality rates describe incident events. Prevalence measures are widely recognized as important for planning

Fig. 3. **Health gaps based on local life expectancy.** The first graph indicates a population with life expectancy at birth of 25 years, while the second graph indicates a population with life expectancy of 37.5 years. In each graph, the bold diagonal represents a survivorship curve. The diagonal to the right of the survivorship curve indicates a survivorship norm in each population that is based on actual life expectancy at each age.



current curative and rehabilitative services, while incidence-based measures are more relevant to the planning of prevention activities. These long-standing arguments have their merits but do not answer the question as to what it means for one population to be healthier than another at a given point in time. It seems undeniable that the common dichotomy between incidence-based and prevalence-based measures does not reflect the composite judgement that an individual behind a veil of ignorance would make on which population is healthier.

As one possible solution to this dilemma we could estimate cohort health expectancy at each age  $x$ , which would reflect both incidence and prevalence.<sup>c</sup> Thus the health expectancy of 50-year-olds depends on the current prevalence of conditions among individuals in this age cohort as well as on the current and future incidence, remission and mortality rates facing the cohort. A period health expectancy at each age  $x$  which reflects incidence and prevalence could also be constructed to provide a measure based only on currently measurable aspects of health. A summary measure for the population could then be based on some aggregation of the cohort or period health expectancies at each age, such as a simple average across all individuals in the population. This aggregate measure would reflect both incidence and prevalence and would be dependent on age structure. Although the mechanics of constructing this measure require further development, we believe that it offers a potential solution to the problem posed jointly by criteria 2 and 3.

**Criterion 4.** If age-specific remission for some health state worse than ideal health is higher, *everything else being equal*, a summary measure should be better. The argument for this criterion is essentially identical to the argument for criterion 3.

<sup>c</sup> Health capital at age  $x$ , proposed as a measure of a cohort's health (53, 54), is a subjective discounted cohort health expectancy at age  $x$ . While it has not been proposed as a summary measure of population health, it includes both prevalence and subjective expected incidence in its arguments.

**Criterion 5.** If two populations A and B include individuals in identically matched health states except for one individual who is in a worse health state in population B, *everything else being equal*, then a summary measure should be worse in B. Here we refer to health states that are described completely in all domains. Any reduction or improvement in any domain would define a new health state. In practice, measurement instruments assign individuals into a finite number of discrete health states in which there is still heterogeneity of levels in each domain of health. At the extreme, measures that categorize the population into only two states, e.g. disabled and not disabled, are insensitive to substantial changes in the true health state of individuals. This criterion is particularly important for assessing the performance of health systems where much of the health expenditure in high-income countries may be directed to interventions that improve the health states of individuals without changing mortality. DFLE, impairment-free life expectancy and dementia-free life expectancy, which all use arbitrary dichotomous weights, do not fulfil criterion 5.

### Other desirable properties of summary measures

Summary measures for comparative purposes are meant to inform many policy discussions and debates. The intended widespread use of summary measures leads to several desirable properties in addition to the basic criteria described above. The appeal of these properties is not based on formal or informal arguments about whether one population is healthier than another but rather on practical considerations such as the following.

- Summary measures should be comprehensible and feasible to calculate for many populations. It is of little value to develop summary measures that will not be used to inform the health policy process. The nearly universal use of a very complex abstract measure, namely period life expectancy at birth, demonstrates that comprehensibility and complexity are different. The interest of the popular press in DALE (82, 83), probably because health expectancies build on life expectancy, is one indication of the comprehensibility of health expectancies. Health gaps are perhaps less familiar to many but the concept is relatively simple and communicable.
- Period-specific summary measures should not change if incidence, remission, prevalence, severity and mortality do not change in the period concerned. In other words, summary measures of the health of a population in a specific period should not depend on the particular set of past events that have preceded the period, nor on the particular set of future events that might follow. This is an important property, as the practical implementation of summary measures of population health by various national health statistics

organizations should be possible using only information available in a given year. Some have argued that period summary measures should include information on the prognoses of individuals in different health states. However, this approach is problematic for a number of reasons. Firstly, as period measures are intended, by definition, to reflect only levels and rates in a defined period, it is logically inconsistent to embed predictions about future events in them. Secondly, if the summary measure is to be used in monitoring change in population health over time, the changes in each individual's health state over time is already reflected in the calculation of the summary measures for each subsequent period. To include predictions about these future changes as part of the summary measure in the current period would be to count these changes twice. Thirdly, the inclusion of prognosis might lead to the rather unappealing conclusion that comparisons of the health of different populations in one period must take account of an infinite stream of events continuing into the future.

- It would be convenient if summary measures were linear aggregates of the summary measures calculated for any arbitrary partitioning of subgroups. Many decision-makers, and very often the public, desire information characterized by this type of additive decomposition across subpopulations. They would like to know what fraction of the summary measure is related to health events in the poor, the uninsured, the elderly, children and so on. Additive decomposition, which also often has appeal for cause attribution, can be achieved for health gaps but not for health expectancies. For example, we can report the number of DALYs in a population for ages 0 to 4 years and for ages 5 years and above, and the sum of these two numbers equals the total health gap in the population. On the other hand it is not clear how to combine the DALE for everybody aged 0 to 4 years with that for everybody aged 5 years and older to obtain a meaningful number. Techniques for estimating the contribution of changes in age-specific mortality rates to a change in life expectancy have been developed (e.g. 84) but they do not have the property of additive decomposition.

### Calculating the contribution of diseases, injuries and risk factors to summary measures

Another fundamental goal in constructing summary measures, which may explain the increasing attention being given to them, is to identify the relative magnitude of different health problems, including diseases, injuries and risk factors, corresponding to uses 5, 6 and 7 above. There are two dominant traditions in widespread use for causal attribution:



categorical attribution and counterfactual analysis. There has been little discussion of their advantages and disadvantages or of the inconsistency of using both approaches in the same analysis. An example of the latter is provided by the Global Burden of Disease 1990 Study (20). Burden attributable to diseases and injuries has been estimated using categorical attribution whereas burden attributable to risk factors or diseases such as diabetes, which act as risk factors, has been estimated using counterfactual analysis.

### Categorical attribution

An event such as death or the onset of a particular health state can be attributed categorically to one single cause according to a defined set of rules. In cause-of-death tabulations, for example, each death is assigned to a unique cause according to the rules of the International Classification of Diseases (ICD), even in cases of multicausal events. For example, in ICD-10, deaths from tuberculosis in HIV-positive individuals are assigned to HIV. This categorical approach to representing causes is the standard method used in published studies of health gaps such as the Global Burden of Disease 1990 (23).

A classification system is required in order that categorical attribution may work. Such a system has two key components: a set of mutually exclusive and collectively exhaustive categories and a set of rules for assigning events to them. ICD, relating to diseases and injuries, has been developed and refined over nearly 100 years. No classification system has been developed for other types of causes such as physiological, proximal or distal risk factors.

### Counterfactual analysis

The contribution of a disease, injury or risk factor can be estimated by comparing the current level and future levels of a summary measure of population health with the levels that would be expected under some alternative hypothetical scenario, for instance a counterfactual distribution of risk or the extent of a disease or injury. The models used in counterfactual analysis may be extremely simple or, in the case of some risk factors with complex time and distributional characteristics, quite complex. The validity of the estimate depends on that of the model used to predict the counterfactual scenarios. Various types of counterfactuals may be used for this type of assessment.

- The effect of small changes in the disease, injury or risk factor can be assessed and the results expressed as the elasticity of the summary measure with respect to changes in the disease, injury or risk factor, or as a numerical approximation of the partial derivative of the summary measure (85, 86).
- Another form of counterfactual analysis assesses the change in a summary measure expected with complete elimination of a disease or injury. A number of studies have presented results on cause-deleted health expectancies (87–91). Wolfson

(92) calculated attribute-deleted health expectancies (i.e. deleting types of disabilities rather than causes).

- More generally, Murray & Lopez (93) have developed a classification of various counterfactual risk distributions that can be used for these purposes, including the theoretical minimum risk, the plausible minimum risk, the feasible minimum risk and the cost-effective minimum risk. The examples of tobacco and alcohol have been used to explore the implications of using these different types of counterfactual distribution to define attributable burden and avoidable burden.
- In intervention analysis the change in a summary measure from the application of a specific intervention is estimated.

Counterfactual analysis of summary measures has a wide spectrum of uses, from the assessment of specific policies or actions to more general assessments of the contribution of diseases, injuries or risk factors. Two important factors are independent of the type of counterfactual used: the duration of the counterfactual and the time during which changes in population health under the counterfactual are evaluated. The complexity of defining the duration of the counterfactual and the time during which change is evaluated can be illustrated with tobacco. A counterfactual for tobacco consumption in which the population does not smoke for one year, followed by a return to the status quo at the end of this year, could be traced out in terms of changes in future health expectancies or future health gaps. Because of time lags and threshold effects, removing a hazard for such a short duration may lead to little or no change in a summary measure of population health over time. Alternatively, the counterfactual change in tobacco could be longer, such as a permanent change to a state of no tobacco consumption.

Models have been developed (94–96) that facilitate counterfactual analyses with varying durations. Changes in the counterfactual distribution of exposure in a population may have an impact on summary measures of population health over many years in the future. Logically, all changes in future population health should be included. For reasons that are debated elsewhere one could argue that changes in the distant future should be weighted as less important than more proximal changes, i.e. future changes should be discounted. One method that has been used is to apply a dichotomous discount rate, such that changes up to time  $t$  are counted equally and changes after time  $t$  are given zero weight. Although discounting is controversial (97–100), choices on the duration of a counterfactual are linked intimately to whether future changes in population health are discounted. For example, a permanent shift in exposure could lead to an infinite stream of future changes in health expectancies or health gaps in the absence of discounting.

As part of its work on comparative risk assessment, WHO is trying to facilitate a debate on

the standardized definition of counterfactuals and the duration of evaluation for a counterfactual change in exposure.

### Advantages and disadvantages

There are three possible ways of analysing the contribution of diseases, injuries or risk factors (see Table 1). Population health can be summarized using health expectancies and health gaps, and cause attribution for diseases and injuries can be assessed using categorical attribution or counterfactuals. Because there is no classification system for risk factors they can only be assessed using the counterfactual approach. Even for diseases and injuries it is not possible to use categorical attribution with a health expectancy, as positive health cannot be assigned to specific diseases or injuries. What are the advantages and disadvantages of the three options in the table?

The advantage of categorical attribution is that it is simple, widely understood and appealing to many users of this information because the total level of the summary measure equals the sum of the contributions of a set of mutually exclusive causes (i.e. categorical attribution produces additive decomposition across causes). The disadvantage is well illustrated by multicausal events such as a myocardial infarction in a diabetic, or liver cancer resulting from chronic hepatitis B. If additive decomposition is a critical property the contribution of diseases and injuries can only be assessed using health gaps.

The counterfactual method for calculating the contribution of diseases, injuries and risk factors has different advantages. It is conceptually clearer, solves problems of multicausality and is consistent with the approach for evaluating the benefits of health interventions (92).

How can causal attribution be used to inform debates on research and development priorities, the selection of national health priorities for action, and health curriculum development? It can be argued that a method of causal attribution should give an ordinal ranking of causes which is identical to the that of the absolute number of years of healthy life gained by a population through cause elimination (or appropriate counterfactual change for a risk factor). This means that the absolute numbers attributable to a cause are

important where cause decomposition is intended to inform public health prioritization.

### Discussion

In this paper we have put forward a basic framework for characterizing and evaluating different types of summary measure. In choosing summary measures for a range of different uses it is critical not only to understand the important differences between the various types of available summary measures but also to distinguish clearly between the range of measures and the different types of instruments and data that may be used as inputs for estimating them. We have defined five basic criteria that may be used as a starting point in evaluating summary measures. We hope that they will provoke further debate on other possible criteria that may be useful to analysts and policy-makers in choosing summary measures for policy applications.

It is worth noting a few examples where even our basic criteria lead to the rejection of certain methodological approaches. For example, the calculation of health gaps using local life expectancy (e.g. 9, 45) violates criterion 1, which requires that as mortality declines a summary measure should improve. According to criterion 5, we should also reject measures that are based on categorizing individuals into two health states, e.g. disabled and not disabled, with arbitrary zero and one weights as in DFLE, ALE and dementia-free life expectancy. A number of remaining health expectancies and health gaps fulfil four of the five criteria, but no measure fulfils the prevalence and incidence criteria at the same time. For comparative uses it may be necessary to develop a new class of measure that reflects both prevalence and incidence, as with the average age-specific health expectancy described above. It is very important to recognize that, for other uses of summary measures, different criteria may be formulated with different implications for the design of such measures.

Causal attribution is a key aspect of summary measures for several important uses outlined above. For diseases and injuries, ICD allows a choice between categorical attribution and counterfactual analysis. The desirability of additive decomposition strongly favours the use of categorical attribution. However, the magnitudes from counterfactual analysis have a more direct and theoretically cogent interpretation. We suggest that in practice the only solution to this tension is routine reporting of both categorical attribution and counterfactual analysis for diseases and injuries. All issues of multicausal death, as with diabetes mellitus, would be well captured in counterfactual analysis even if categorical attribution tends to underestimate the problem. There is no classification system for risk factors, whether physiological, proximal or distal, and consequently the only option is counterfactual analysis. There are many options for defining counterfactuals, and

Table 1. Approaches to analysing the contributions of diseases, injuries or risk factors to summary measures of population health

	Categorical attribution	Counterfactual analysis
Health expectancies		<ul style="list-style-type: none"> <li>• Diseases</li> <li>• Injuries</li> <li>• Risk factors</li> </ul>
Health gaps	<ul style="list-style-type: none"> <li>• Diseases</li> <li>• Injuries</li> </ul>	<ul style="list-style-type: none"> <li>• Diseases</li> <li>• Injuries</li> <li>• Risk factors</li> </ul>

substantial work is needed to understand more fully the implications of adopting different approaches.

Improving the estimation of summary measures of population health depends on designing the most appropriate measures for particular purposes. It also requires improvement of the empirical basis for the epidemiology of fatal and non-fatal health outcomes, including attribution by cause, and for health state valuations. One critical requirement is an improved understanding of the determinants of differences between self-reported and observed measures of performance or capacity in selected domains of health.

In proposing this framework for choosing summary measures we have invoked both a common-sense notion whereby, in some cases, everybody could agree that one population was healthier than another, as well as a more formal mechanism for defining this choice, using Harsanyi's notion of choice behind a veil of ignorance. There are some potentially important implications of the veil of ignorance framework for choosing a summary measure of population health for comparative purposes. For example, the current methods used to measure preferences for time spent in health states may not be entirely consistent with this framework,

and modified methods would perhaps need to be developed. Clearly, it would be helpful to provide a more rigorous formal treatment of this approach.

As work on summary measures gathers speed, their uses and complexities are becoming more widely appreciated. The application of simple criteria may lead to the rejection of some measures and the development of new ones. An extensive developmental agenda exists; nevertheless, the use of summary measures should not be delayed until all methodological issues have been resolved. Every effort should be made to use currently available summary measures that satisfy as many of the criteria and desirable properties as possible. The calculation of alternative summary measures should be facilitated by making the critical information on the epidemiology of non-fatal health outcomes and mortality widely available. ■

### Acknowledgements

The authors are grateful for valuable comments by Frances Kamm, John Broome, Julio Frenk, Rafael Lozano, Alan Lopez, Dan Wikler, Jan Barendregt, Sarah Marchand, Paul van der Maas and Sander Greenland.

## Résumé

### Examen critique des mesures synthétiques de l'état de santé d'une population

Au cours des dix dernières années, on s'est de plus en plus intéressé à l'élaboration, au calcul et à l'utilisation des mesures synthétiques de l'état de santé d'une population, associant les informations sur la mortalité et les pathologies non mortelles. Le présent article passe en revue les questions fondamentales et les difficultés rencontrées pour concevoir et appliquer ce type de mesures ; il propose également un cadre pour évaluer les différentes possibilités.

La conception d'une mesure synthétique peut dépendre de son utilisation. Plusieurs usages peuvent en être faits : par exemple, comparaison de l'état de santé des populations entre différents pays ou dans le même pays sur plusieurs périodes, ou encore évaluation de la part relative des maladies, traumatismes ou facteurs de risque dans la charge totale de morbidité pour une population. A cause de ces usages destinés à influencer sur la politique, les aspects normatifs des mesures synthétiques doivent être examinés avec soin.

Le vaste éventail des mesures synthétiques proposées peut être divisé en deux grandes catégories : celles ayant trait à l'espérance de santé, qui étend la notion d'espérance de vie au temps passé dans un état de santé qui n'est pas optimal, et celles concernant les lacunes de santé, qui mesurent la différence entre l'état de santé d'une population et la norme établie. Il existe un certain nombre d'exigences et de questions communes à toutes les mesures synthétiques. Les principales données à recueillir comprennent les renseignements sur la mortalité en fonction de l'âge, l'épidémiologie des affections non mortelles et aussi la valeur attachée aux différents états de santé par rapport à ce que l'on considère comme étant la santé idéale ou par

rapport à la mort. On peut calculer ces mesures pour une période ou une cohorte donnée du point de vue de la prévalence ou de l'incidence. La définition et la mesure de l'état de santé d'une population peuvent varier considérablement, et il existe un certain nombre de problèmes conceptuels et méthodologiques en suspens concernant la valeur à accorder aux différents états de santé.

Dans le présent article, nous proposons un ensemble de critères de base et de caractéristiques souhaitables susceptibles d'être utilisés pour distinguer les différentes mesures synthétiques. L'application de ces critères de base pourrait conduire à rejeter certaines mesures et à en élaborer de nouvelles. Par exemple, les lacunes de santé qu'on définit par rapport à l'espérance de vie locale sont en contradiction avec le critère fondamental selon lequel, lorsque la mortalité décroît, tout autre paramètre étant équivalent par ailleurs, une mesure synthétique doit s'améliorer. Nous concluons également qu'aucune mesure synthétique actuelle ne répond simultanément aux deux critères voulant que ces mesures s'améliorent lorsque soit l'incidence, soit la prévalence d'un état pathologique diminue.

Il reste beaucoup à faire pour poursuivre la mise au point des mesures synthétiques ; toutefois, il ne faut pas attendre d'avoir résolu tous les problèmes méthodologiques pour les utiliser. Pour les appliquer actuellement, il faut se concentrer sur celles qui répondent à autant de critères de base et de caractéristiques souhaitables que possible. Les efforts pour améliorer la base empirique de l'épidémiologie des pathologies mortelles et non mortelles et de la valeur attribuée aux différents états de santé doivent se poursuivre.

## Resumen

### Análisis crítico de los indicadores sinópticos de la salud de la población

En la última década ha crecido el interés por el desarrollo, cálculo y uso de indicadores sinópticos de la salud de la población, que combinan información sobre la mortalidad y sobre las consecuencias de los problemas de salud no mortales. En este artículo se analizan los aspectos y desafíos más importantes del diseño y aplicación de indicadores sinópticos, y se presenta un marco de evaluación de las diferentes alternativas.

Las características de un indicador sinóptico dependerán del uso previsto. Se pueden emplear con diversos fines, por ejemplo para comparar la salud de la población en diferentes países o en el mismo país a lo largo del tiempo, o para evaluar en qué medida contribuyen relativamente las diferentes enfermedades, traumatismos y factores de riesgo a la carga de morbilidad total de una población. Los aspectos normativos de los indicadores sinópticos deben estudiarse detenidamente, ya que esos usos previstos pueden influir en las políticas.

La amplia gama de indicadores sinópticos propuestos puede dividirse en dos grandes familias: las esperanzas de salud, que amplían el concepto de esperanza de vida para tener en cuenta el tiempo transcurrido en estados de salud distintos de la perfecta salud; y las desigualdades en salud, que reflejan la diferencia entre la salud en una población y una norma establecida para la salud de la población. Hay una serie de requisitos de información y otros aspectos comunes a todos los indicadores sinópticos de la salud de la población. Entre las aportaciones más importantes destacan la información sobre la mortalidad por edades y la epidemiología de los resultados sanitarios sin consecuencias mortales, y los valores atribuidos a los

diferentes estados de salud en relación con la salud ideal o con la muerte. Los indicadores sinópticos se pueden calcular con métodos basados en periodos o en cohortes, y a partir tanto de incidencias como de prevalencias. El estado de salud de una población se puede definir y cuantificar de muy diversas maneras, y su evaluación plantea una serie de problemas conceptuales y metodológicos relevantes.

En el presente artículo proponemos una serie de criterios básicos y condiciones deseables para distinguir los diferentes indicadores sinópticos. La aplicación de estos criterios básicos puede llevar a rechazar algunos de esos indicadores y a desarrollar otros. Por ejemplo, las desigualdades en salud definidas con respecto a la esperanza de vida local no cumplen el criterio básico en virtud del cual, si la mortalidad disminuye, *ceteris paribus*, el indicador sinóptico debe mejorar. También llegamos a la conclusión de que ninguno de los indicadores sinópticos actuales satisface a la vez los dos criterios que exigen que mejore tanto si disminuye la incidencia como si disminuye la prevalencia del problema de salud en cuestión.

Queda aún mucho trabajo por hacer para perfeccionar los indicadores sinópticos, pero no debemos esperar a que se hayan resuelto todos los problemas metodológicos para hacer uso de ellos. Habría que centrarse fundamentalmente en los que satisfacen el mayor número posible de criterios básicos y de propiedades deseables. Deben proseguir los esfuerzos a fin de mejorar la base empírica de la epidemiología de los resultados sanitarios mortales y no mortales y la evaluación de los diferentes estados de salud.

## References

1. **Field MJ, Gold GM, eds.** *Summarizing population health: directions for the development and application of population metrics.* Washington DC, National Academy Press, 1998.
2. **Sanders BS.** Measuring community health levels. *American Journal of Public Health*, 1964, **54**: 1063–1070.
3. **Chiang CL.** *An index of health: mathematical models* (United States Public Health Services Publications Series 1000, Vital and Health Statistics Series 2, No. 5). Washington DC, National Center for Health Statistics, 1965.
4. **Sullivan DF.** *Conceptual problems in developing an index of health* (United States Public Health Services Publications Series 1000, Vital and Health Statistics Series 2, No. 17). Washington DC, National Center for Health Statistics, 1966.
5. **Sullivan DF.** A single index of mortality and morbidity. *HSMHA Health Reports*, 1971, **86**: 347–354.
6. **Piot M, Sundaresan TK.** *A linear programme decision model for tuberculosis control. Progress report on the first test-runs.* Geneva, World Health Organization, 1967 (unpublished document VMO/TB/Tech.Information/67.55).
7. **Fanshel S, Bush JW.** A health-status index and its application to health services outcomes. *Operations Research*, 1970, **18** (6): 1021–1066.
8. **Katz S et al.** Measuring the health status of populations. In: Berg RL, ed. *Health status indices.* Chicago, Hospital Research and Educational Trust, 1973: 39–52.
9. **Ghana Health Assessment Project Team.** A quantitative method of assessing the health impact of different diseases in less developed countries. *International Journal of Epidemiology*, 1981, **10** (1): 72–80.
10. **Preston SH.** Health indices as a guide to health sector planning: a demographic critique. In: Gribble JN, Preston SR, eds. *The epidemiological transition. Policy and planning implications for developing countries.* Washington DC, National Academic Press, 1993.
11. **Robine J-M et al., eds.** *Calculation of health expectancies: harmonization, consensus achieved and future perspectives.* London, John Libbey Eurotext, 1993.
12. **Mathers C, McCallum J, Robine J-M, eds.** *Advances in health expectancies: Proceedings of the 7th Meeting of the International Network on Health Expectancy (REVES).* Canberra, Australian Institute of Welfare, 1994.
13. **Katz S et al.** Active life expectancy. *The New England Journal of Medicine*, 1983, **309** (20): 1218–1224.
14. **Bone NIR.** International efforts to measure health expectancy. *Journal of Epidemiology and Community Health*, 1992, **46**: 555–558.
15. **Bronnum-Hansen H.** Trends in health expectancy in Denmark, 1987–1994. *Danish Medical Bulletin*, 1998, **45** (2): 217–221.

16. **Crimmins EM, Saito Y, Ingegneri D.** Trends in disability-free life expectancy in the United States, 1970–90. *Population and Development Review*, 1997, **23** (3): 555–572.
17. **Mutafova M et al.** Health expectancy calculations: a novel approach to studying population health in Bulgaria. *Bulletin of the World Health Organization*, 1997, **75** (2): 147–153.
18. **Sihvonen AP et al.** Socioeconomic inequalities in health expectancy in Finland and Norway in the late 1980s. *Social Science and Medicine*, 1998, **47**: 303–315.
19. **Valkonen T, Sihvonen AP, Lahelma E.** Health expectancy by level of education in Finland. *Social Science and Medicine*, 1997, **44**: 801–808.
20. **Murray CJL, Lopez A, eds.** *The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries and risk factors in 1990 and projected to 2020*. Cambridge, MA, Harvard School of Public Health on behalf of the World Health Organization and The World Bank, 1996 (Global Burden of Disease and Injury Series, Vol. 1).
21. **Murray CJL, Lopez A.** *Global health statistics*. Cambridge, MA, Harvard School of Public Health on behalf of the World Health Organization and The World Bank, 1996 (Global Burden of Disease and Injury Series, Vol. II).
22. **Murray CJL, Lopez A.** Evidence-based health policy – lessons from the Global Burden of Disease Study. *Science*, 1996, **274**: 740–743.
23. **Murray CJL, Lopez A.** Mortality by cause for eight regions of the world: Global Burden of Disease Study. *Lancet*, 1997, **349**: 1269–1276.
24. **Murray CJL, Lopez A.** Regional patterns of disability-free life expectancy and disability-adjusted life expectancy: Global Burden of Disease Study. *Lancet*, 1997, **349**: 1347–1352.
25. **Murray CJL, Lopez A.** Global mortality, disability, and the contribution of risk factors: Global Burden of Disease Study. *Lancet*, 1997, **349**: 1436–1442.
26. **Murray CJL, Lopez A.** Alternative projections of mortality and disability by cause 1990–2020: Global Burden of Disease Study. *Lancet*, 1997, **349**: 1498–1504.
27. **Lozano R, Frenk J, Gonzalez MA.** El peso de la enfermedad en adultos mayores, México 1994. [The burden of disease in older adults, Mexico 1994.] *Salud Pública de México*, 1994, **38**: 419–429.
28. **Lozano R et al.** Burden of disease assessment and health system reform: results of a study in Mexico. *Journal for International Development*, 1995, **7** (3): 555–564.
29. *Health and the economy: proposals for progress in the Mexican health system*. Mexico, Fundación Mexicana para la Salud, 1995.
30. *La carga de la enfermedad en Colombia*. [The burden of disease in Colombia.] Bogota, Ministry of Health, 1994.
31. **Ruwaard D, Kramers PGN.** *Public health status and forecasts*. The Hague, National Institute of Public Health and Environmental Protection, 1998.
32. **Bowie C et al.** Estimating the burden of disease in an English region. *Journal of Public Health Medicine*, 1997, **19**: 87–92.
33. **Concha M et al.** *Estudio carga de enfermedad: informe final*. [Burden of disease study, final report.] Estudio Prioridades de Inversión en Salud Ministerio de Salud, 1996.
34. **Murray CJL, Acharya AK.** Understanding DALYs. *Journal of Health Economics*, 1997, **16**: 703–730.
35. **Murray CJL et al.** *US patterns of mortality by county and race: 1965–1994*. Cambridge, Harvard Center for Population and Development Studies and Centers for Disease Control, 1998.
36. **Bobadilla JL.** *Searching for essential health services in low- and middle-income countries: a review of recent studies on health priorities*. Washington DC, The World Bank, 1996.
37. **Brundtland, GH.** *The global burden of disease*. (Speech delivered on 15 December 1998, Geneva.)
38. *The world health report 1999 – making a difference*. Geneva, World Health Organization, 1999.
39. **Murray CJL, Frenk J.** A framework for assessing the performance of health systems. *Bulletin of the World Health Organization*, 2000, **78** (6): 717–731.
40. **Gakidou E, Murray CJL, Frenk J.** Defining and measuring health inequality: an approach based on the distribution of health expectancy. *Bulletin of the World Health Organization*, 2000, **78** (1): 42–54.
41. **Murray CJL, Gakidou E, Frenk J.** Health inequalities and social group differences: what should we measure? *Bulletin of the World Health Organization*, 1999, **77** (7): 537–543.
42. *Investing in health research and development. Report of the Ad Hoc Committee on Health Research Relating to Future Intervention Options*. Geneva, World Health Organization, 1996 (document TDR/Gen/96.1).
43. **Brock DW.** Ethical issues in the development of summary measures of population health status. In Field MJ, Gold GM, eds. *Summarizing population health: directions for the development and application of population metrics*. Washington DC, National Academy Press, 1998.
44. **Murray CJL.** Rethinking DALYs. In: Murray CJL, Lopez A, eds. *The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries and risk factors in 1990 and projected to 2020*. Cambridge, MA, Harvard School of Public Health on behalf of the World Health Organization and The World Bank, 1996 (Global Burden of Disease and Injury Series, Vol. 1).
45. **Williams A.** Calculating the global burden of disease: time for a strategic reappraisal? *Health Economics*, 1999, **8**: 1–8.
46. **Anand S, Hanson K.** DALYs: efficiency versus equity. *World Development*, 1998, **26** (2): 307–310.
47. **Daniels N.** Distributive justice and the use of summary measures of population health status. In Field MJ, Gold GM, eds. *Summarizing population health: directions for the development and application of population metrics*. Washington DC, National Academy Press, 1998.
48. **Williams A.** Intergenerational equity. An exploration of the ‘fair innings’ argument. *Health Economics*, 1997 **6** (2): 117–132.
49. **Ware JE, Sherbourne CD.** The MOS 36-item short-form health survey (SF-36). 1. Conceptual framework and item selection. *Medical Care*, 1992, **30**: 473–483.
50. **Erickson P, Wilson R, Shannon E.** *Years of healthy life*. Hyattsville, Maryland; US National Center for Health Statistics, 1995.
51. **Wilkins R, Adams OB.** Quality-adjusted life expectancy: weighting of expected years in each state of health. In: Robine J-M, Blanchet M, Dowd JE, eds. *Health expectancy* London, HMSO, 1992 (OPCS studies on medical and population subjects No. 54).
52. **Ritchie K et al.** Dementia-free life expectancy: preliminary calculations for France and the United Kingdom. In: Robine J-M et al., eds. *Calculation of health expectancies: harmonization, consensus achieved and future perspectives*. London, John Libbey Eurotex, 1993.
53. **Cutler DM, Richardson E.** Measuring the health of the US population. *Brookings Paper: Microeconomics*, 1997: 217–227.
54. **Cutler DM, Richardson E.** The value of health. *American Economic Review*, 1998, **88** (2): 97–100.
55. **Dempsey M.** Decline in tuberculosis: the death rate fails to tell the entire story. *American Review of Tuberculosis*, 1947, **56**: 157–164.
56. **Romedor JM, McWhinnie JR.** Potential years of life lost between ages 1 and 70: an indicator of premature mortality for health planning. *International Journal of Epidemiology*, 1977, **6**: 143–151.
57. **Hyder AA, Rotllant G, Morrow RH.** Measuring the burden of disease: healthy life-years. *American Journal of Public Health*, 1998, **88** (2): 196–202.
58. **Shryock HS, Siegel JS.** *The methods and materials of demography: the life table*. Washington DC, US Bureau of the Census, October 1971: 443–446.
59. **Longnecker NT.** The Framingham results on alcohol and breast cancer. *American Journal of Epidemiology*, 1999, **149** (2): 93–101.

60. **Deeg DJH, Kriegsman DMW, van Zonneveld RJ.** Trends in fatal chronic diseases and disability in the Netherlands 1956–1993. In: Mathers C, McCallum J, Robine J-M, eds. *Advances in health expectancies: Proceedings of the 7th Meeting of the International Network on Health Expectancy (REVES)*. Canberra, Australian Institute of Health and Welfare, 1994: 80–95.
61. **Barendregt JJ, Bonneux L.** Changes in incidence and survival of cardiovascular disease and their impact on disease prevalence and health expectancy. In: Mathers C, McCallum J, Robine J-M, eds. *Advances in health expectancies: Proceedings of the 7th Meeting of the International Network on Health Expectancy (REVES)*. Canberra, Australian Institute of Health and Welfare, 1994: 345–354.
62. **Wolfbein S.** The length of working life. *Population Studies*, 1949, **3**: 286–294.
63. **Branch LG et al.** Active life expectancy for 10,000 Caucasian men and women in three communities. *Journal of Gerontology*, 1991, **46** (4): M145–150.
64. **Rogers A, Rogers RG, Branch LG.** A multistate analysis of active life expectancy. *Public Health Reports*, 1989, **104** (3): 222–226.
65. **Mathers CD, Robine J-M.** How good is Sullivan's method for monitoring changes in population health expectancies? *Journal of Epidemiology and Community Health*, 1997, **51**: 80–86.
66. **Robine J-M, Blanchet M, Dowd EJ, eds.** *Health Expectancy*. London, HMSO, 1992.
67. **Barendregt JJ, Bonneux L, Van der Maas PJ.** Health expectancy: an indicator for change? *Epidemiology of Human Health*, 1994, **48**: 482–487.
68. **Mathers CD, Robine J-M.** How good is Sullivan's method for monitoring changes in population health expectancies? Reply. *Journal of Epidemiology and Community Health*, 1997, **51**: 578–579.
69. **McDowell I, Newell C.** *Measuring health: a guide to rating scales and questionnaires*. Second Edition. New York, Oxford University Press, 1996.
70. **Bickenbach JE et al.** Models of disablement, universalism and the international classification of impairments, disabilities and handicaps. *Social Science and Medicine*, 1999, **48**: 1173–1187.
71. **Murray CJ, Chen LC.** In search of a contemporary theory for understanding mortality change. *Social Science and Medicine*, 1993, **36** (2): 143–155.
72. **Erickson P.** Evaluation of a population-based measure of quality of life: the health and activity limitation index (HALex). *Quality of Life Research*, 1998, **7**: 101–114.
73. **EuroQol Group.** EuroQol: a new facility for the measurement of health-related quality of life. *Health Policy*, 1990, **16**: 199–208.
74. **Mehrez A, Gafni A.** Quality-adjusted life-years, utility theory and healthy-year equivalents. *Medical Decision Making*, 1989, **9**: 142–149.
75. **Johannesson M, Pliskin JS, Weinstein MC.** Are healthy-years equivalents an improvement over quality-adjusted life years? *Medical Decision Making*, 1993, **13** (4): 281–286.
76. **Cuyler AJ, Wagstaff A.** QALYs versus HYEes. *Journal of health economics*, 1993, **11**: 311–323.
77. **Nord E.** The person trade-off approach to valuing health care programs. *Medical Decision Making*, 1995, **15** (3): 201–208.
78. **Nord E et al.** Incorporating societal concerns for fairness in numerical valuations of health programmes. *Health Economics*, 1999, **8**: 25–39.
79. **Harsanyi JC.** Cardinal utility in welfare economics and in the theory of risk-taking. *Journal of Political Economy*, 1953, **61**: 434–435.
80. **Rawls J.** *A theory of justice*. Cambridge, Harvard University Press, 1971.
81. **Broome J.** *Ethics out of economics*. New York, Cambridge University Press, 1999.
82. **Crosette B.** Americans enjoy 70 healthy years, behind Europe, UN says. *New York Times*, 5 June 2000: A10.
83. **Brown D.** New look at longevity offers disease insight. *Washington Post*, 12 June 2000: A9.
84. **Arriaga EE.** Measuring and explaining the change in life expectancies. *Demography*, 1984, **21**(1): 83–96.
85. **Hill GB, Forbes WF, Wilkins R.** *The entropy of health and disease: dementia in Canada*. 9th meeting of the International Network on Health Expectancy (REVES), Rome, 11–13 December 1996.
86. **Mathers CD.** Gains in health expectancy from the elimination of diseases among older people. *Disability and Rehabilitation*, 1999, **21** (5-6): 211–21.
87. **Colvez A, Blanchet M.** Potential gains in life expectancy free of disability: a tool for health planning. *International Journal of Epidemiology*, 1983, **12**: 224–229.
88. **Mathers CD.** *Estimating gains in health expectancy due to elimination of specified diseases*. Fifth Meeting of the International Network on Health Expectancy (REVES), Ottawa, 19–21 February 1992.
89. **Mathers CD.** *Gains in health expectancy from the elimination of disease: a useful measure of the burden of disease?* Tenth Meeting of the International Network on Health Expectancy (REVES), Tokyo, 9–11 October 1997.
90. **Nusselder WJ.** *Compression or expansion of morbidity? A life-table approach* [Ph.D. Thesis]. Rotterdam, Erasmus University, 1998.
91. **Nusselder WJ et al.** The elimination of selected chronic diseases in a population: the compression and expansion of morbidity. *American Journal of Public Health*, 1996, **86** (2): 187–193.
92. **Wolfson MC.** Health-adjusted life expectancy. *Health Reports*, 1996, **8** (1): 41–45.
93. **Murray CJL, Lopez A.** On the comparable quantification of health risks: lessons from the Global Burden of Disease Study. *Epidemiology*, 1999, **10**(5): 594–605.
94. **Gunning-Schepers LJ.** The health benefits of prevention: a simulation approach. *Health Policy*, 1989, **12** (1–2): 35–48.
95. **Gunning-Schepers LJ, Barendregt JJ, van der Maas PJ.** Population interventions reassessed. *Lancet*, 1989, **1**(8636): 479–481.
96. **Gunning-Schepers LJ, Barendregt JJ.** Timeless epidemiology or history cannot be ignored. *Journal of Clinical Epidemiology*, 1992, **45** (4): 365–372.
97. **Cropper MI, Aydede SK, Portney PR.** Preferences for life saving programs: how the public discounts time and age. *Journal of Risk and Uncertainty*, 1994, **8**: 243–265.
98. **Redelmeier DA, Heller DM.** Time preference in medical decision making and cost-effectiveness analysis. *Medical Decision Making*, 1993, **13**: 505–510.
99. **Viscusi VK, Moore MJ.** Rates of time preference and valuations of the duration of life. *Journal of Public Economics*, **38**: 297–317.
100. **Sen AK.** Isolation, assurance and the social rate of discount. *Quarterly Journal of Economics*, 1967, **81**: 1123–1124.