Distr.
GENERAL

CES/AC.36/1998/29
EUR/ICP/INFO 020603/29

5 October 1998

ENGLISH AND FRENCH ONLY

STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE

WORLD HEALTH ORGANIZATION
REGIONAL OFFICE FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Joint ECE-WHO Meeting on Health Statistics
(Rome, Italy, 14-16 October 1998)

SESSION III: Development and use of health output indicators

MEASURING HEALTH - VISIONS AND PRACTICALITIES

Paper submitted by Statistics Canada*

Introduction

A major challenge in health statistics is achieving a reorientation in emphasis. At present, the preponderance of countries' "health" data pertain to the health care system, and then mainly its inputs and throughputs. There is a paucity of statistical information on the levels, trends and distribution of health status for countries' populations, and for various relevant sub-groups. We know far more about the costs of health care, and the numbers of patients treated, than we do about the health impacts of the treatments, and the health of the population in general.

The main exceptions in terms of broadly available statistics draw on mortality data, particularly to produce figures on infant mortality and life expectancy. But these data also shed, at best, very indirect light on the health status among the living.

By way of introduction, this paper begins with a few paragraphs of Canadian background relating to the measurement of population health status. The main part of the paper then reviews the leading approaches to the

---

* Prepared by Michael C. Wolfson. This paper represents the personal views of the author, who is responsible for any errors or infelicities.

measurement of population health, and discusses several conceptual and ethical challenges.

Canadian Background

Almost a decade ago, Statistics Canada undertook an internal review of its health statistics program (Wolfson, 1991). This review concluded that there were two major problems, but they did not pertain to the data that were being collected, which were generally of very good quality. Rather, one problem was that the various data streams being collected did not fit together; there was a lack of coherence. The other major problem was a data gap. There was a fundamental imbalance -- the data being collected mainly covered the health care sector, and then mostly its inputs and throughputs. There were almost no data on health outcomes (nor on health-related factors outside the health care system).

Also about a decade age, the Canadian Institute for Advanced Research (CIAR) initiated a Program in Population Health. This group of researchers adopted a broad perspective regarding the determinants of health and developed an innovative synthesis of the research literature (Evans et. al., 1994). Their analysis has had a major impact on Canadian thinking regarding health and health care, including the requirements for Canada's health information system.

Subsequently, a milestone in this area was the report of the National Task Force on Health Information (Wilk, 1991). It built substantially on the kinds of insights generated by the CIAR's Population Health Program. It concluded, among other things, that Canada needed highly multivariate person-oriented longitudinal microdata (to help disentangle the more complex determinants of health), and more importantly for the analysis in this paper, a focal measure of health outcomes for the Canadian population - "The health information system should include an overall aggregate index of population health -- some sort of GDP or CPI of health, which would be the culmination or aggregation of a coherent family of health status indicators." The National Forum on Health (1997) reinforced this observation in its recent recommendations.

Largely in response to these developments, Statistics Canada was funded to begin the National Population Health Survey, which began collecting data in 1994 (Catlin and Will, 1992). This is a longitudinal survey, conducted every two years. It includes data on individual health status, as well as informed consent to link the survey responses to the respondents' health care records. In parallel, an analytical group within Statistics Canada has been developing summary measures of population health status.

Main Approaches to Measuring Population Health Status

There is a wide variety of statistics and statistical measures used to indicate the health status of populations such as that of a country and its main population groups. Still, any such statistics must start with data

describing the health status of the individuals within the population, what we can refer to as the individual-level health status **descriptive system.** This set of descriptions may be derived from a population survey, or administrative data such as the computerized portions of hospital discharge abstracts. In either case, the starting point is descriptions of the health status for each in a group of individuals who are representative of the population.

There are several approaches to health state descriptions. One important distinction is whether a **clinical** or a **vernacular** description is used. The former is typically produced by a health care professional, usually a physician, in the course of a diagnosis, and is expressed most often using the WHO's ICD. In Canada, for example, these data are readily available for all in-patient hospital visits (about 4 million per year). Somewhat similar though not as detailed data are available for a very large portion of physician encounters as a result of the coding required of fee-for-service physicians for them to obtain reimbursement.

The other main source of data on individuals' health status is from sample surveys of the individuals themselves. Since they are not expert in clinical diagnosis, these surveys typically ask for vernacular descriptions of health status, such as pain, impairments and disabilities, and limitations in ordinary activities of daily living (ADLs). These tend not to be *ad hoc* questions; rather, disabilities and ADLs are elicited using standardized questionnaires that have proven reliable and meaningful on prior surveys. In some cases, the question sets are the product of international consensus, such as the disability questions developed by the OECD. In other cases, they are proprietary, such as SF-36, originally developed originally by the RAND corporation.

The second main step in constructing statistical measures of population health is aggregation. Some method is applied to the large set of individual health state descriptions, of whatever kind, in order to derive a handful of summary statistics at the population level. We refer to this as population- or **macro-level aggregation.** The simplest and by far the most widely used method is cross tabulation, typically generating statistics like the proportion of the population (or a sub-group defined for example by age and sex) suffering from health problem X.

However, this straightforward cross-tabular approach becomes unwieldy when a number of different conditions are being monitored, and one wants to make comparisons over a number of years, or across a range of population sub-groups, or between countries. One is then faced with a combinatorial explosion in the numbers of statistics.

Another concern is that the (sub-)populations being compared may differ in systematic ways which are already well understood. In such cases, it

would be better to make comparisons after "taking out" such systematic forms of variation. The most widely recognized form of such systematic variation is age structure, and the usual approach is some sort of age-standardization. Mechanically, this corresponds to re-weighting the individual-level health status descriptive data, before cross-tabulating, in order to represent some "standard" or reference population, for example in terms of its distribution by age groups.

More generally, this process of standardization can usefully be thought of as the simulation of a "counter-factual" population. It involves constructing the sample or population of individual health status descriptions that *would* exist if, say, it had the same age distribution or other set of characteristics as a given reference population. Once this hypothetical or counter-factual population has been constructed, aggregation by cross-tabulation (if working directly with micro data) or re-weighting (if working with meso-level or partially aggregated data) can then proceed in the usual way. Age-standardized mortality rates are perhaps the best-known example, typically constructed by re-weighting meso-level age-specific mortality rates.

In order to reduce the combinatoric explosion of possible summary statistics, two other broad forms of aggregation have been developed. One is at the level of the individual health status descriptive system. These descriptive systems generally produce a profile for each individual - for example the levels of an individual's functioning on a range of dimensions such as pain, mobility, sensory perception and cognition. At this individual level, responses to such a series of questions are aggregated to generate, for each individual, some sort of summary health status score or index - what we refer to as **micro-level** aggregation. For example, there are depression scales that aggregate responses to a series of specific questions, and more importantly overall summary measures such as the McMaster University Health Utility Index (Torrance, 1986; Feeney et. al., 1995), and the Euro-Qol measure (Dolan et. al., 1994) that aggregate across a carefully specified set of individual-level health domains and generate an index, usually in the zero to one interval. There is considerable methodological work underway in this area, particularly regarding the ways to assign a number in the zero-to-one interval to a person-year summarizing that individual's health status (e.g. Gold et. al., 1996, Institute of Medicine, 1998).

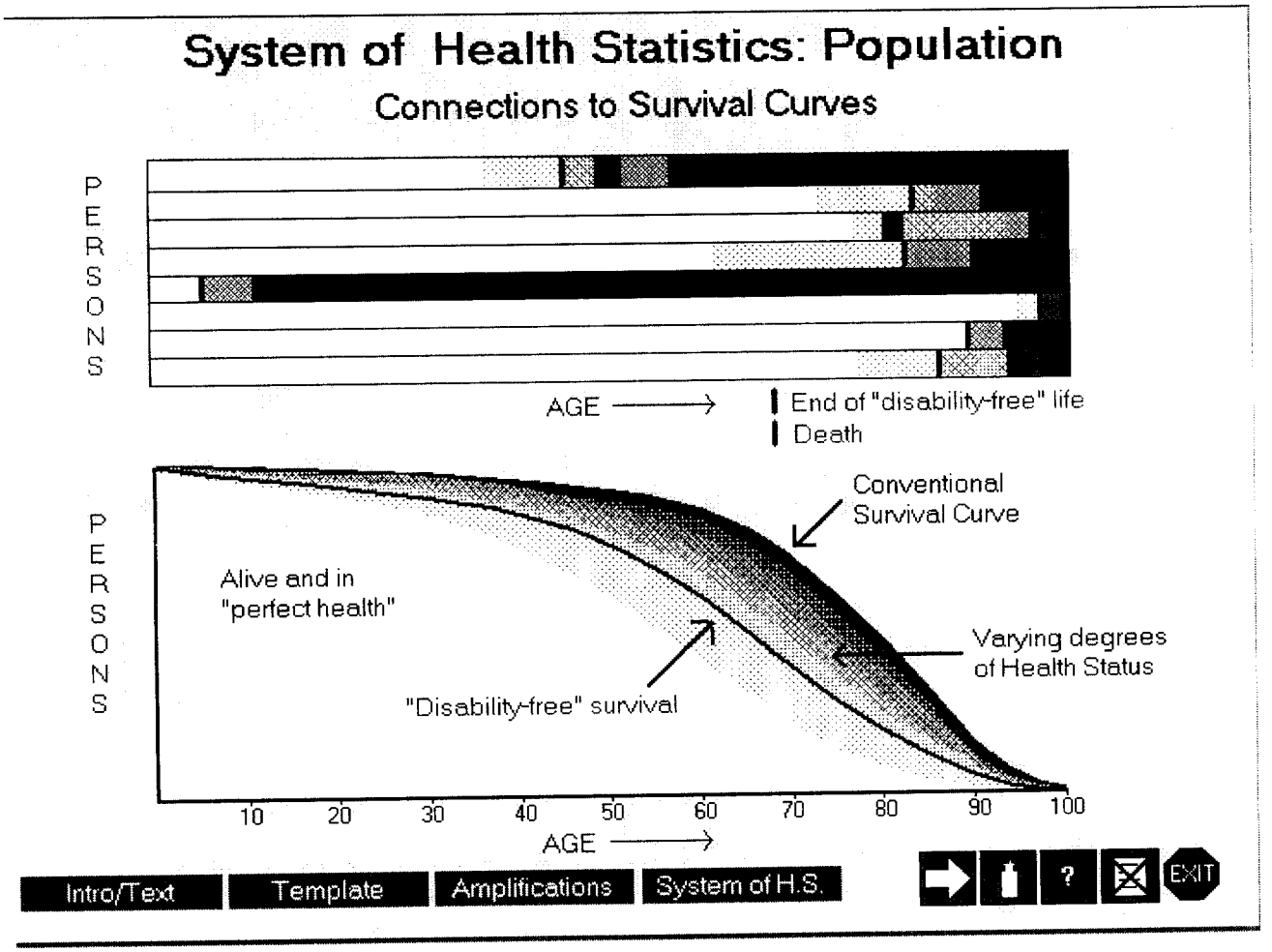
It is then straightforward to (macro-) aggregate the resulting individual-level summary index scores to the population level via usual cross-tabular methods, for example to generate average levels of individual health status by sex and age groups. This kind of two step aggregation, first at the individual (micro) level, and then to the population (macro) level, is essential for the derivation of summary indicators of health status at the population level. Still, for interested users, it should always be possible to access the underlying population survey (or other source) micro data so that the "reasons" for any interesting aggregate patterns (e.g. the

main factors contributing to the declining health status of the elderly) can be retrieved. In other words, a "drill down" capability should be a part of any reasonable statistical system generating summary measures of population health status.

The Concept of Integrated Measures of Population Health Status

The other broad form of aggregation involves alternatives to cross-tabulation when aggregating over individuals up to the population level, i.e. alternative **macro**-aggregation approaches. The best known examples here are variants of life expectancy, such as disability-free life expectancy (DFLE; Mathers and Robine, 1993), disability-adjusted life expectancy (DALE; Murray, 1996), and health-adjusted life expectancy (HALE; Mathers, Robine, and Wilkins, 1994; OECD, 1998; Roberge et al., 1997). These measures have the advantage, as summary population health indicators, of combining information on health status among the living, and mortality rates.

Figure 1, drawn from a conceptual framework for health information (Wolfson, 1992) developed by the National Task Force on Health Information (Wilk, 1991), illustrates the core ideas underlying these summary measures of population health status. The diagram starts in the top half with a symbolic representation of a handful of individual's life paths - shown as horizontal bars. Rather than a simple alive-dead dichotomy, each individual's life course trajectory of health status is represented by levels of gray shading. White indicates perfectly healthy, black = dead (with a short red mark denoting the point of death), and the various grays in between show intermediate health states. At the same time, it has been assumed that what ever multi-dimensional individual-level health status descriptive system has been used, some sort of micro-level aggregation has been applied so that each moment along an individual's life path can be reasonably characterized by a single index of health status taking values between zero or black (dead) and one or white (fully healthy).



The next step in the process is shown in the bottom half of the diagram. Conventionally, survival curves are derived from a relatively simple life table, using age-specific mortality rates. However, survival curves can also be built up microanalytically from a large number of individual life paths such as those shown at the top of the figure. The process is very straightforward. Imagine that we have a much greater number of the kinds of life paths or individual health biographies shown at the top, indeed a sample representative of the population. They need only be sorted by age at death, and then stacked up, in order to derive the overall survival curve shown in the bottom half of the diagram. The area under this curve is then the very familiar and widely accepted indicator, life expectancy.

But this survival curve by itself conveys nothing at all regarding health status among the living. However, we have, in this case, built up the survival curve by aggregating a representative sample of individual life paths where each path also includes a sequence of micro-level summary health index scores. It is therefore possible to shade in the area beneath the survival curve, according the amounts of time each individual in the underlying sample spent at various levels of health status.

Alternatively, it is possible to select a threshold level of health status or functioning (often based on a different set of questions than those used to compute the summary micro-level index of health) below which we can label the individual as "disabled" (denoted by a short blue mark in the individual life paths in the top portion of the diagram). Then, just as we sorted and stacked an imagined representative sample of individual life paths to construct the survival curve, we could sort the individual life paths in order of the age at which the disability threshold was reached, and then stack them to obtain a "disability-free" survival curve. This too is shown in the bottom portion of the diagram.

We can now see that macro summary measures of population health status like DFLE, DALE, and HALE are all similar, since they are all based of the shaded survival curve shown in the bottom of the diagram. DFLE is simply the area under the "disability-free" survival curve. Alternatively, DFLE can be thought of as a special case of HALE where the micro-level aggregation is based on a valuation function such that disability states less severe than the threshold are assigned a value of 1 (i.e. the same value as "fully healthy"), while those that are more severe than the threshold are assigned a value of 0 (i.e. the same value as "dead")

HALE is similar, but uses a more "responsive" weighting or health status valuation. It is the area under the overall survival curve weighted by the darkness of the shading. DALE (Murray, 1996) is a variant of HALE where processes of expert consensus were used to assign numerical scores representing a uni-dimensional severity of disability, directly tied to various clinical disease states, rather than deriving the scores via micro-level aggregation over individual level health profiles based on an explicit health status descriptive system.[1]

Illustration of Health-Adjusted Life Expectancy

Members of the HALE family of summary indicators represent the prime candidates for measures of population health status. These indicators can be constructed not only for the entire population, but also for population subgroups. Figure 2 provides examples for the Manitoba population broken down into sub-groups based on two alternative socio-economic status (SES) variables, educational attainment and family income (Nault, Roberge, and Berthelot, 1996). One set of results shown in the graphs is conventional

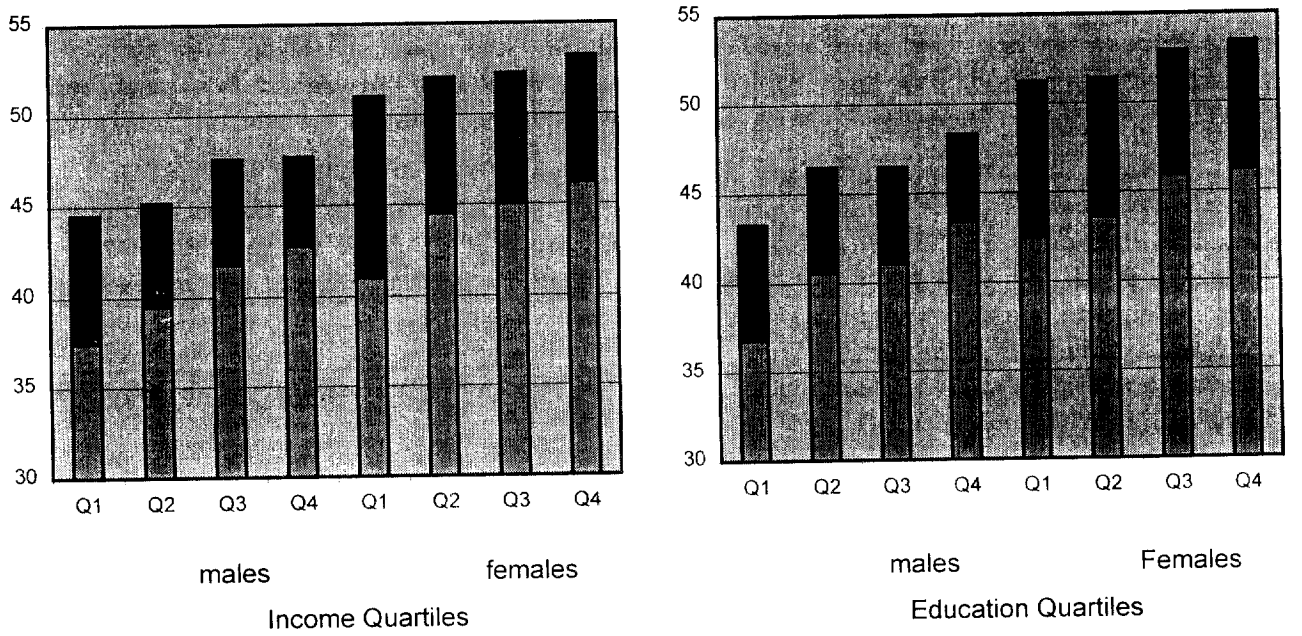
---

[1] The more widely known DALYs highlighted in the Global Burden of Disease work (World Bank, 1993) are more complex than DALEs since they also typically include some sort of age-weighting, a discount factor, and an element akin to Potential Years of Life Lost (PYLL), the gap in life expectancy relative to that of a reference population.

life expectancy (LE), and the other is health-adjusted life expectancy (HALE).

For both males and females, considerable differences across SES quartiles are clearly evident, with individuals in the highest education and income categories living longer, and in better health than those in the lowest categories. Also clearly evident for both the income and education groupings is a gradient. In other words, the relationship is not a threshold phenomenon - steps up the socio-economic scale, even starting further up the scale, are generally associated with an increment in health status, measured either as LE or as HALE.

Figure 2: Life Expectancy (both bars) and Health-Adjusted Life Expectancy (lower bar)at age 30 Using Health Status Data for 1994



The magnitudes of the differences in HALE between high and low quartiles is 3.7 to 6.6 years for income and quartiles is 3.7 to 6.6 years for income and education, and men and women. These are wider than for LE, which differs by amounts ranging from 2.2 to 5.0 years between the high and low quartiles. Thus, estimates of SES differences in mortality may understate the magnitude of health inequalities when population health status is measured more broadly to encompass morbidity as well as mortality. Figure 2 gives only a one-time set of estimates. However, if we had a time series of these HALE and LE estimates, we would also be able to shed light on one of the grand debates in epidemiology - the so-called "compression of morbidity" - whether or not increasing life expectancy is associated with a greater or lesser proportion of the life course spent in poor health.

Illustration of "Cause-Deleted" HALE

The most widely used method for constructing HALE and related summary measures of population health status is the Sullivan method (Sullivan, 1971). It draws on two kinds of data. One is the mortality rates used to construct conventional life tables and life expectancy estimates. The other is a population survey which is first micro-aggregated and then cross-tabulated by sex and age group (e.g. five years) to generate average levels of health status. (Alternatively, responses to other health status questions can be similarly cross-tabulated tc generate proportions of the population "free of disability", resulting in the DFLE measure.) Then these two sets of partially aggregated or "meso-"data are combined by multiplying (1) the number of person-years lived in each age group from the life table by (2) the average health status of that same age group (or the proportion non-disabled) observed in the health survey. After this modification to the year-by-year (or more typically quinquennial) life table results, "health adjusted" (or disability-free) life expectancy is computed in the standard manner.

Alternatively, and in line with the description of Figure 1, HALE can be estimated microanalytically - by describing the life paths, and more specifically the health status trajectories, of a longitudinal sample of individuals. Statistics Canada has been developing a POpulation HEalth Model (POHEM; Wolfson, 1994) one of whose objectives is precisely the capacity to estimate a representative sample of such life paths including their health status trajectories. Table 1 illustrates this capacity with estimates that generalize the widely used notion of cause-deleted life expectancy (LE).

Two sorts of generalization are illustrated. One goes beyond LE to include HALE as well. In other words, the impact of "deleting a cause" of ill health is estimated both in terms or mortality and in terms of morbidity, as measured by a summary measure of individual health status. The other goes beyond the usual ICD definitions of "cause". Typically, cause-deleted life expectancy is straightforwardly estimated by starting with all cause mortality rates (by age and sex), then subtracting the portion of each mortality rate "caused by" a given disease such as lung cancer, and finally recomputing life expectancy in the usual manner. With POHEM, we can go beyond this by examining underlying or "upstream" causes as well, such as smoking.

Table 1 -- Cause-Deleted LE and HALE Estimates

	females		males	
Scenario	LE	HALE	LE	HALE
base	79.9	74.9	74.0	70.6
no lung cancer	+0.8	+0.6	+0.9	+0.7
no smoking	+0.6	+0.5	+1.0	+0.9
heavy smoking	-1.2	-1.1	-2.1	-2.0
no arthritis	nil	+2.0	nil	+1.0

Several points regarding these estimates are worth noting. First, like typical cause-deleted LE figures, these are based on period life tables. In other words, they are the answers to questions of the following form: what if a cohort was born in Canada in the early 1990s, and throughout its life experienced the transition probabilities observed in the early 1990s, except that health problem X were not present in the population at all? This is clearly a hypothetical birth cohort, not only because of the complete absence of health problem X, but also because the cohort is not living in real calendar time. Rather it is assumed to live in some imaginary world where all the dynamics of the early 1990s are frozen - the relevant mortality rates, risk factor prevalences, risk functions and disease progression rates are assumed always to have been the same, and to persist indefinitely into the future.

While these may appear to be strong assumptions, they are fundamental to the most widely used indicator of population health status, LE, as well as widely used variants like cause-deleted LE. The generalizations in Table 1 should also make clear that LE and its related measures are the products of simulation. POHEM builds on this core notion and generalizes it. It also makes the notion of cause-deletion potentially more reasonable. For example, it is very difficult to imagine the elimination of lung cancer without the elimination of smoking. But if smoking were eliminated, heart disease incidence and fatality (among other "tobacco-sensitive diseases) would also decline and occur at later ages. Thus, lung cancer-deleted life expectancy embodies a very naïve underlying causal story.

POHEM, on the other hand, provides a framework within which much more realistic causal stories can be incorporated. In the case of the estimates in Table 1, the elimination of smoking is assumed to affect both lung cancer and coronary heart disease incidence. However, while these diseases represent the two largest pathways by which smoking affects health, they still represent only about half of all tobacco-sensitive causes of death. (POHEM is still under development, and models for other tobacco-sensitive diseases have not yet been implemented.) As a result, the numbers for smoking in Table 1 should not be taken as anything more than illustrative;

they likely under-estimate the effects of smoking by a factor of two. The figures for (osteo-) arthritis, however, are relatively complete.

Table 1 shows an overall life expectancy for females of 79.9 years, with health-adjusted life expectancy about 5.0 years less at 74.9 (base case). Hypothetically eliminating lung cancer increases males' life expectancy by 0.9 years, but increases their health-adjusted life expectancy by 0.2 years less, since those additional 0.9 years are not all spent in full health. Alternatively, eliminating all smoking increases males' life expectancy by about 1.0 years (again bearing in mind that the only effects taken into account are via lung cancer and heart disease). Hypothetically eliminating arthritis has no effect on life expectancy, but has three or four times the effect on HALE for females as eliminating lung cancer or smoking.

Since osteo-arthritis is almost never a cause of death in mortality data, it never ranks on the "league tables" of importance of diseases, for example used to motivate charitable giving or to allocate health research monies. However, its burden in terms of morbidity would put it in the top handful of diseases. And smoking is never shown as a cause of death, yet it is more important a cause of morbidity and mortality than lung cancer. These scenarios, particularly eliminating arthritis versus eliminating smoking or lung cancer for females, therefore suggest that this kind of extension of HALE indicators may have an important influence on our basic sense of health priorities. Moreover encompassing the health burdens of both diseases and risk factors, within a common analytical framework, and measured in terms of both mortality and morbidity, provides a highly useful and coherent family of summary population health status indicators.

A Vision for Population Health Status Measurement

1. Summary Health Status Indices

Given the discussion so far of methodology for summary measures of population health status, and the illustrations, we turn now to a "vision" for such measures. Ideally, we would like a coherent and integrated statistical framework, with a summary measure of population health status at 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 would provide a broad population-based overview of trends and patterns. Thus, the overall summary measure, such as HALE, would give a time series not only for the population as a whole, but would also be replicated for relevant population sub-groups (e.g. as shown in Figure 2 above) to facilitate within and between group analysis.

These macro summary measures would be based on:

(1) a standardized health status descriptive system at the individual or micro-level,

(2) a micro-level aggregation or valuation function that would map the various levels of health status along each dimension of the micro-level descriptive system into the interval from zero to one, and

(3) the macro-level aggregation process implicit in the HALE measure.

2. A Coherent and Powerful System of Health Statistics

Moreover, the micro-level descriptive system with its associated valuation function (i.e. items 1 and 2 just above) would provide a common metric not only for monitoring population health status (e.g. via regular population health surveys), but also for evaluating health benefits across interventions and generating consistent evidence across observational studies. In other words, whatever other health outcome and covariate data were collected in randomized clinical trials (RCTs), in population-based epidemiological studies (e.g. longitudinal cohort studies), and possibly in carefully specified sub-sets or samples of otherwise routine administrative data for the health care system such as hospital discharge abstracts, data for the same standardized descriptive system would also be collected.

The use of a common metric in RCTs, in cohort studies, and in selected administrative data streams would be a major advance, since it would allow meaningful comparisons across studies and across the gamut of routine health care. At present, the only common end-point in RCTs and cohort studies is mortality. But with the increasing prevalence of chronic disease morbidity, this is clearly inadequate for contemporary health, health care, and health policy analysis. Some sort of standardized measure(s) of morbidity is essential, as recommended by Gold et. al. (1996). This would aid the accumulation of research evidence, e.g. via meta analyses; it is key to the comparative evaluation of health interventions; and it is central to the routine monitoring of the outcomes of the health care system.

Such a common metric for individual-level health status is also necessary for "meta synthesis" - for weaving together data from diverse sources, for example as needed in the construction of models like POHEM. The cause-deleted HALE results generated by POHEM simulations, and shown in Table 2 above, are only possible if data from diverse surveys, from the health care administrative data, from RCTs, and from cohort studies can be woven together.

Finally, using this same individual-level metric in constructing the aggregate population index as well requires that it is regularly collected for a representative population sample. In turn, this baseline representative sample would provide an important context for micro level studies. It would allow the results of RCTs or cohort studies to be judged in terms of an overall population measure of the relative burdens of various health problems. This in turn would greatly extend the capacity for embedding cost-effectiveness evaluations of health interventions in a "burden

of disease" context - thereby allowing existing or prospective interventions to be judged both in terms of cost-effectiveness, and their relative impacts in reducing disease and ill health.

Moreover if, as with Canada's National Population Health Survey, the survey is both longitudinal and can be exactly matched to individual data on encounters with the health care system, it is possible to develop much stronger and ongoing evaluations of the real "outputs" of health care - the actual improvements in population health being generated. This is in line with growing interest in regular "report cards" on the functioning and performance of the health care sector.

In sum, a central part of the vision for a coherent and powerful system of health statistics is the widespread adoption of a common metric for individual-level health status - across household surveys used for monitoring population health status, across epidemiological cohort studies designed to elicit new information on long run causal relationships, and across RCTs designed to assess the relative efficacy of alternative health interventions.

3. A System of Health Statistics Linked to Policy-Relevant Questions

This system of health statistics, with the framework for measuring health status at both the individual and population levels just described, would represent a major advance in our ability to monitor population health (both levels and distributions), and to accumulate knowledge about causal factors. There is a further need, however, to connect this information to public policy. The "bottom line" is informing collective decisions on which of the myriad known health-affecting interventions merit a share of society's limited resources.

Thus, at the macro level, an obvious desideratum is to be able to link the summary measure of population health status back to factors amenable to policy. At one level, this is an unreasonable expectation. For example, no one seriously questions the production of data on unemployment or income inequality just because there is no simple "policy lever" to which each responds. It is recognized that both are influenced by a variety of factors, only some of which are under the control of governments. At the same time, data on a large part of the range of influential factors (subject to budgetary constraints for the statistical system, and limits in knowledge) are routinely collected in most countries (e.g. educational attainment, productivity, income tax liabilities). And finance and treasury ministries have for decades constructed complex models to enable policy analysts to simulate the likely impacts of moving one or other of the available policy levers (e.g. income tax provisions).

An analogous approach is required in the area of health policy. The starting point is agreement on the target or objective, namely population health status (both level and distribution). This would be routinely indicated by a macro HALE summary measure forming the apex of a coherent

framework of related health indicators. Methodologically, this is straightforward if the HALE measure is built up from underlying life paths as indicated in Figure 1. It can then be decomposed into any number of components such as sojourn times (e.g. durations of individuals' lifetimes spent in one or another health state) like DFLE, and breakdowns by population groups.

The HALE family of indicators can also be extended by explicitly embedding it in a simulation model like POHEM. This would broaden the family of related measures in ways analogous to cause-deleted life expectancy as shown in Table 1 above. A related generalization would be "population attributable fractions" - for example recent estimates that "40,000 deaths in Canada are due to smoking" would become "x years of HALE are lost due to the pain and mobility impairment aspects of arthritis". Recall that these "cause-deleted" and "attributable fraction" analogues are effectively answers to "what if" questions. They therefore require credible statistical representation of the causal pathways that are implicated in answering the "what if" questions, as well as a sophisticated simulation modeling capacity to generate plausible versions of the implied hypothetical or "counter-factual' scenarios.

This is feasible, as demonstrated by Statistics Canada's POHEM model, but ideally requires an unprecedented degree of planning and coordination in the collection of data, and highly skilled analytical staff.

Main Conceptual and Ethical Challenges

The vision just outlined for a coherent and powerful system of health statistics, with a summary measure of population health status at the apex, raises significant conceptual and ethical challenges.

Individual-Level Health Status Descriptive System

At the foundation is the desirability of using the same micro summary health status measure for both a range of research applications (e.g. RCTs and cohort studies) and for population level health status information. However, there is no a consensus (in Canada or elsewhere, though the U.S. is moving in this direction; see Gold et. al., 1996 and Institute of Medicine, 1998) that the specific micro-level measure of individual health status should be broadly promoted. Nor is there consensus that a macro HALE measure should be featured prominently in various national compendia of health statistics, though there is widespread interest (e.g. OECD, 1998).

The hesitancy turns on a several key questions. One is how to select and operationalize the implicit definition of health. This includes the number and specific kinds of dimensions (e.g. mobility, dexterity, vision, pain), and the number and character of the levels of functioning or health status along each of these dimensions (ranging from very modest to very severe health problems).

My view is that we should confine this concept to functional limitations expressed in the vernacular - "disabilities" in the ICIDH lexicon. Clearly, this omits both clinical disease and social role functions from the core concept, but it seems better to consider these as precursors and sequalae respectively. The omission of social role function, as an ultimate outcome, is troubling. But the focus on functional limitations ("within the skin" health characteristics) offers several advantages - it is more amenable to validation; it may prove a less complex task from the viewpoint of achieving the consensus needed for widespread adoption; and it is more likely to allow structurally independent dimensions of health status in the individual-level descriptive system, a technical benefit for the derivation of the valuation function (see below). The omission of clinical disease can be accommodated by appropriate data sets bridging the clinical and vernacular - specially developed "Rosetta Stone" micro data sets, such as exactly matched population health survey and clinical administrative data..

A related challenge is achieving international comparability. It would be ideal if an international body like the OECD or WHO could promulgate a consensus on the description of individual-level health status, much as the ILO has developed a consensus on the measurement of unemployment. If such a consensus descriptive system were routinely used in countries' household health surveys, research could benefit from the natural experiments offered by different countries' experiences. Also, the required international consensus process would give a degree of legitimacy to the results.

The Micro-Level Valuation Function

A further set of challenges arise in the choice of the individual-level or micro valuation function. On what ethical foundation should one base a single method to weight and then aggregate disjoint aspects of myriad individuals' health such as mobility and pain? One approach is to draw a representative sample of the population and then, using a carefully constructed set of questions, elicit each person's implicit valuation function. For the most part, such valuation functions have been elicited in terms of individual preferences. However, given their most likely uses, it seems more appropriate to elicit values for social trade-offs (as argued in Nord et. al., 1993; Nord, 1998).

In either case (individual or social valuations), the specific questions involved in eliciting valuations tend to be cognitively complex; and for many respondents they pose difficult ethical questions. Thus, it seems better to precede eliciting individuals' valuations by group discussion, as an adaptation of deliberative polling (e.g. Murray, 1996).

In empirical work to date, there is also evidence of a lack of robustness to rather subtle differences in the way the valuation questions are framed. Moreover, it seems that health state valuation functions vary systematically across sub-groups -- e.g. rich and poor; and it is well known that they vary with personal or close experience of disease or disability.

Nevertheless, just as virtually all statistical offices worldwide use an *average* expenditure basket for constructing consumer price indices and then measuring inflation, even though there are important differences i expenditure patterns across population sub-groups, it should be satisfactory to use an "average" valuation function, at least as a starting point.

It may be of practical importance to realize that progress on consensus for an individual-level health status descriptive system can be de-coupled from that for a micro-level valuation function. An interim valuation function can always be used, as in Canada's National Population Health Survey results for 1994 and 1996. Then, (assuming no change in the descriptive system) revised health status indicators can be tabulated or constructed once a "final" valuation function is agreed.

Macro-Aggregation

A further set of conceptual and ethical challenges lies in the method used to aggregate individual health status (based on a micro-level health status descriptive system and a micro-level valuation function) into a macro summary population health status index. The central challenge is the way judgements regarding distributional equity are incorporated, particularly insofar as they are used in cost-effectiveness evaluations to determine who can be given which kinds of treatment or health care. In effect, the HALE summary measure gives "one person-year, one vote". An increment of 0.1 (say) in individual health status for one year increases HALE by exactly the same amount no matter who the person is - age, sex, income and health status trajectory make no difference.

This may be problematic, as argued by Brock (1998) and by Daniels (1998). For example, they give examples where a matter of indifference in the calculation of HALE (or equivalence classes in more technical parlance), would not be matters of indifference to the individuals concerned (e.g. whether or not they or someone else received the treatment), or to certain population sub-groups (e.g. the disabled). There are two main responses to this set of concerns. One is that many of their challenges to the equivalence classes implicit in a given HALE measure reduce to challenges to the individual-level valuation function being used. However, if this valuation function has been elicited from a representative population using a social valuation perspective (such as Nord's person-trade-off), they are in effect saying they disagree with the values expressed by a representative sample of the population. To this they are entitled, but it does not seem a very strong philosophical argument.

Another of their challenges concerns population sub-groups who, given the characteristics of their group (e.g. disabled), would be treated differently if resources were allocated solely based on the macro index. One response is to appeal to the Rawlsian notion of the "veil of ignorance". The fundamental question is whether an individual would say that a given pattern of health intervention financing is fair - prior to knowing which specific

health problems might afflict them over their lifetime. This seems a better criterion of the fairness of a system of social allocations of resources for health-related interventions than one that polls individuals after they know that they have disease x or health problem y.

The other response to this challenge of population sub-groups with specific health problems is to recall that what is being proposed is not HALE in isolation. Rather, HALE is envisaged as the apex of a coherent system of health statistics. In particular, this system should include the capability to "drill down" and estimate the impacts of a given intervention on various sub-groups. Thus, a capacity to generate measures of the distributional impacts of health-related interventions is very much part of the proposed statistical system. This is exactly analogous to the measures of the distributional impacts of income tax changes that are taken for granted by finance ministry analysts, their ministers, and the general public. It is also an extension of the long-standing advice (though not always followed in practice) of cost-benefit analysis more generally - where any project that is expected to have both significant gainers and losers, it should be evaluated not only on the basis of its overall cost-benefit ratio, but also its net benefits (and losses) for various sub-groups.

Concluding Comment

There is no fully scientific basis for answering all these challenges. In the end, processes of informed consensus building are needed. Moreover, we should not set the bar too high in terms of conceptual clarity and ethical accord. A reasonable standard of quality here is the System of National Accounts and its prominent summary index, GDP. There is a long tradition in welfare economics to show that the ethical foundations for measuring the national income are far from robust. Yet the SNA has long played a major and useful role in international and domestic economic policy. Similarly, an unattainable perfection should not block the obvious utility of a coherent and powerful system of health statistics based on standardized measures of health status.

References

Brock, D.W. (1998), "Ethical Issues in the Development of Summary Measures of Population Health Status", in Institute of Medicine (1998), Summarizing Population Health, Directions for the Development and Application of Population Metrics, National Academy Press, Washington, D.C..

Catlin, G. and B.P. Will (1992), "The National Population Health Survey: Highlights of Initial Developments", Health Reports 4(3):313-319.

Daniels, N. (1998), "Distributive Justice and the Use of Summary Measures of Population Health Status", in Institute of Medicine (1998), Summarizing Population Health, Directions for the Development and Application of Population Metrics, National Academy Press, Washington, D.C.

Dolan, P., C. Gudex, P. Kind, and A. Williams (1994), The Measurement and Valuation of Health. First Report on the Main Survey, The MVH Group, Centre for Health Economics, University of York. May, 1994.

Evans, R.G. et al. (1994), Why are Some People Healthy and Others Not? The Determinants of Health of Populations, Aldine de Gruyter, New York.

Feeny, D., W. Furlong, M.H. Boyle, and G.W. Torrance (1995), "Multi-Attribute Health Status Classification Systems: Health Utilities Index", PharmacoEconomics 7(6):490-502.

Gold, M.R., L.B.Russell, J.E.Siegel, and M.C.Weinstein (1996), Cost Effectiveness in Health and Medicine, Oxford University Press.

Institute of Medicine (1998), Summarizing Population Health, Directions for the Development and Application of Population Metrics, National Academy Press, Washington, D.C.

Mathers, C. and J-M Robine (1993), "Health expectancy indicators: a review of the work of REVES to date", in J-M Robine, C.D.Mathers, M.B.Bone, I.Romieu (Eds), Calculation of Health Expectancies: Harmonization, Consensus Achieved and Future Perspectives, INSERM / John Libby Eurotext Ltd., Vol. 226.

Mathers, C.D., J.-M. Robine, and R. Wilkins (1994), "Health Expectancy Indicators: Recommendations for Terminology", in C.D. Mathers, J. McCallum, and J.-M. Robine (Eds) Advances in Health Expectancies: Proceeding of the 7th Meeting of the International Network on Health Expectancy (REVES), Canberra, February 1994. Australian Institute of Health and Welfare: AGPS, Canberra.

Murray,C.J.L. (1996), "Rethinking DALYs" in Murray,C.J.L., and A.D. Lopez (Eds), The Global Burden of Disease, Vol. 1, World Health Organization, Harvard University Press.

National Forum on Health (1997), Canada Health Action: Building on the Legacy. Final Report of the National Forum on Health, Health Canada, Ottawa.

Nault, F., R. Roberge, and J-M Berthelot (1996), "Esperance de vie et esperence de vie en sante selon le sexes, l'etat matrimoniale el el statut socio-economique au Canada", Cahiers quebecois de demographie, Vol 25, no 2, automne p 241-259.

Nord, E., J. Richardson, and K. Macarounas-Kirchmann (1993), "Social Evaluation of Health Care Versus Personal Evaluation of Health States", International Journal of Technology Assessment in Health Care 9(4):463-478.

Nord, E. (1998), "The Validity of Summary Health Indices", ECE/WHO Meeting on Health Statistics.

OECD (1998), "Summary Statement from the Ad Hoc Meeting of Experts in Health Statistics, 3-5 December, 1997", Paris.

Sullivan, D.F. (1971), "A Single Index of Mortality and Morbidity", HSMHA Health Reports 86(4):347-354.

Torrance, G.W., Measurement of health state utilities for economic appraisal - a review", Journal of Health Economics, 5:1-30.

Wilk, M.B. (1991), Report of the National Task Force on Health Information, National Health Information Council, c/o Statistics Canada, Ottawa.

Wolfson (1991), "A System of health Statistics - Toward a New Conceptual Framework", Review of Income and Wealth, 37(1), p63-80.

Wolfson (1992), "A Template for Health Information", World Health Statistics Quarterly, 45(1), p109-113 (including software diskette).

Wolfson (1994) "POHEM - A Framework for Understanding and Modeling the Health of Human Populations", World Health Statistics Quarterly 47(3):157-176.