PART 2

FOR RESEARCHERS AND STATISTICIANS

SOURCES OF INFORMATION

Any attempt at measuring socioeconomic inequalities in health at the national, regional or local level should start by assessing the availability of data. The assessment process should include an inventory of data available on socioeconomic inequalities in health and then an assessment of the informative value of these data. Finally, if necessary, provisions should be made for generating new data.

Identifying available data on socioeconomic inequalities in health

When data on socioeconomic inequalities in health are being inventoried, it is convenient first to determine which data are available on the frequency of health problems in the general population and then to see which of these data can be categorized by socioeconomic status.

Commonly available data on the frequency of health problems in the population

Throughout Europe, the two main sources of information on the health status of the population are registries and specially designed surveys. Table 1 provides a brief overview of commonly kept registries *(9)*. Registries fall into four groups, depending on the event being registered: vital events (births and deaths), instances of health care utilization, receipt of social security benefits and a miscellaneous group of other events (such as traffic accidents). In contrast to registries, registers link data on the same individuals from different registries to arrive at a more complete or more valid measurement of the incidence or prognosis of a particular disease (for example, cancer registers or mental health registers).

Table 1. Registries: an overview

Type of registry	Examples of information registered
Vital events	Births
	Deaths
Health care utilization	Hospital admissions
	Mentally ill people in institutions
Social security benefits	Sickness absence
	Long-term work disability
Other events	Traffic accidents
	Medical examination of conscripts for military service

Some of these data sources have intrinsic limitations in measuring the magnitude of socioeconomic inequalities in health. First, the inequalities in the frequency of health problems as they occur in the population can only be estimated accurately if there are no differences between socioeconomic groups in the probability of being registered as soon as the health problem occurs. For example, socioeconomic differences in rates of hospital admission only serve as a valid proxy for differences in the prevalence or incidence of the underlying conditions if the probability of being admitted when the condition arises is the same for individuals of higher and lower socioeconomic status. Although the degree of bias varies between countries and possibly between different parts of one country, it is generally preferable to use data that are not affected by this type of problem. Of the registries mentioned in Table 1, only registries of vital events and registers more or less fulfil this condition.

A second limitation of some of these data sources is that they cover only a relatively small part of the total spectrum of the health problems in the population. Registries of traffic accidents and infectious diseases and registers of cancer and mental health problems are examples of very important but not very comprehensive data sources. A more comprehensive view is offered by registries of hospital admissions (with the limitation mentioned previously) and by registries of deaths (since death may be caused by a wide range of different conditions).

It is for these reasons that mortality data are so frequently used in measuring socioeconomic inequalities in health and that mortality data are recommended here as one of the two core elements of any such measurement exercise, especially monitoring changes over time. The continuous nature for collecting mortality data systems makes these especially suitable for measuring trends in socioeconomic inequalities in health.

The second main source of information on the health status of the population is specially designed surveys among samples of the population *(10)*. The two basic options are health examination surveys, in which a nurse or doctor physically examines the individuals in the sample, and health interview surveys in which data collection is completely based on self-reports by the respondents. As health examination surveys are costly, only health interview surveys are used on a larger scale, in various parts of Europe. Data collection procedures may involve a personal interview, a telephone interview or a postal survey or combinations of these.

Health interview surveys consist of sets of questions or questionnaires (question batteries), each of which measures one or more aspects of health or another relevant phenomenon. Because of advances in psychometrics, many of the questionnaires have been properly validated, although some, even frequently used questionnaires, lack a good evaluation *(11,12)*. Table 2 gives an overview of the different types of questions and questionnaires that can be used in health interview surveys. The measurement of the physical, mental and social aspects of health permits a comprehensive picture of the health status of individuals. In practice, many health interview surveys focus on general health and physical aspects of health.

Because of their comprehensiveness and the fact that they complement mortality data, health interview survey data are recommended here as the second core element of any attempt at measuring socioeconomic inequalities in health. The relatively low costs of these surveys make repeated measurements feasible and thereby greatly enhance the possibilities for monitoring trends.

Table 2. Health interview surveys: types of basic questionnaires available

Type of survey	Examples
General health	Perceived health status Nottingham Health Profile Medical Outcomes Study Short-Form General Health Survey (SF-36)
Physical health	Inventory of chronic conditions Indicator on long-term disabilities from the Organisation for Economic Co-operation and Development
Mental health	General Health Questionnaire Zung Self-rating Depression Scale
Social health	Interview Schedule for Social Interaction RAND Social Activities Battery

Categorizing morbidity and mortality by socioeconomic status

The calculation of morbidity and mortality rates among individuals of higher or lower socioeconomic status requires that socioeconomic information be linked to the morbidity or mortality information somewhere in the analysis.

For morbidity data, the situation is relatively straightforward. Health interview surveys aim at measuring the prevalence of certain health problems at one point in time and can therefore be used to estimate the differences in prevalence between socioeconomic groups. The simplest way of doing this is to include a set of questions to determine the socioeconomic characteristics of the respondents in the survey. In technical terms, the analysis of socioeconomic differences in the prevalence of health problems is cross-sectional in character, because the measurement of health problems refers to the same point in time as the measurement of socioeconomic status. Clearly, this leads to problems of inference on the causal nature of the association between socioeconomic status and health as observed in these data: did low socioeconomic status produce bad health or vice versa *(13,14)*? This problem cannot be solved with this type of data, but fortunately the problem is not as serious in monitoring efforts as it is in research projects aiming at establishing causal relationships.

For mortality data, the situation is more complex. Mortality is not measured as a prevalence rate but as an incidence rate: the rate at which death occurs in a certain number of individuals who are observed over a certain period of time *(15–17)*. Here the most appropriate way of categorizing the data by socioeconomic status is to measure everyone's socioeconomic status at the start of the period over which they are followed. This longitudinal approach has been applied successfully in Scandinavian countries where, thanks to the availability of personal identification numbers, population census records and the death registry can be linked, so that the national population can be followed up for mortality after each population census *(18–21)*. France and England and Wales have no personal identification numbers, but a sample of the consensus population can be followed up for mortality by manually linking population records to the death registry *(22,23)*. Longitudinal mortality studies are also available for some specific regions or cities; for example, where an epidemiological study has been carried out among a sample of the population *(24–28)*.

When longitudinal data are not available, sometimes a good alternative exists in the form of linked cross-sectional data. Such data can be obtained from registries that not only record the socioeconomic characteristics of each subject included but, in addition, record the reasons for any withdrawal from the registry (death or otherwise). With such a registry, the socioeconomic status of each person who dies can be determined. The registry of socioeconomic characteristics also supplies the denominator data on the numbers of people, according to socioeconomic status, who were at risk of dying during the period in which the deaths occurred. This approach is called cross-sectional because the numerator and denominator data refer to the same period of time. One of the few examples comes from (western) Germany, where the social insurance and pension registries are the most important source of nationwide data on mortality rates according to occupational status and educational level *(29,30)*.

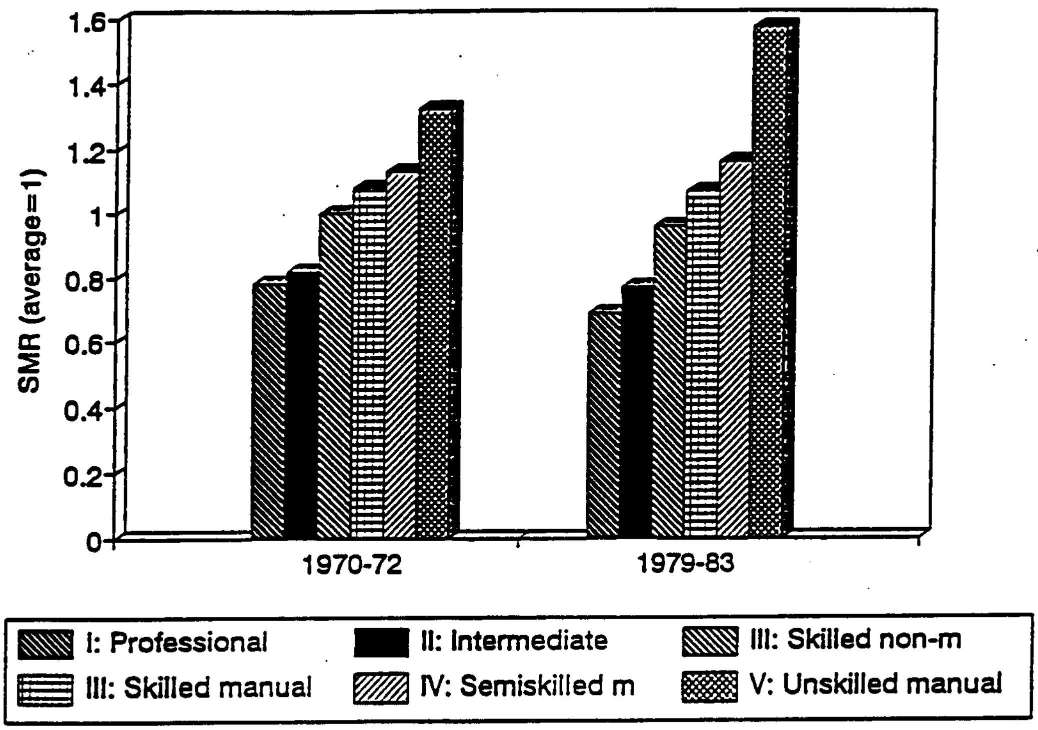
A more widely available but less satisfactory variant of cross-sectional data is unlinked cross-sectional data. Again, the number of deaths observed during a time interval can be related to the number of people who were exposed to the risk of dying during that same time interval. The term unlinked refers to the fact that the socioeconomic information on the deaths is not derived by linkage. Instead, the socioeconomic information on the deaths and on the population at risk comes from two different sources:

- the death registry, which provides the numbers of deaths according to the socioeconomic characteristics registered on the death certificate;
- the population census, which provides the corresponding numbers of people at risk according to the same socioeconomic characteristics determined during the census.

Dividing the socioeconomic status-specific numbers of deaths (the numerator) by the socioeconomic status-specific numbers of person-years (the denominator) provides estimates of mortality rates by socioeconomic status.

The classical example of this type of study is the decennial supplements on occupational mortality of the Registrar General in the United Kingdom *(31,32)*. Fig. 1 shows the variation in standard mortality rates according to social class for two periods around the population censuses of 1971 and 1981. In both periods, lower social classes have higher mortality rates. The largest mortality differences are observed for the most recent period. This indication of an increasing inequality in mortality generated wide interest and stimulated various studies on trends in inequalities in mortality in England and Wales *(33–38)*.

Fig. 1 Trends in mortality differences by social class in England and Wales among men aged 45–64 years



Source: Office of Population Censuses and Surveys *(31,32)*.

Similar studies were carried out in many other countries, including Hungary, Ireland, Spain, Switzerland, and the former Yugoslavia *(39–43)*. In many countries, however, the validity of this type of study may be compromised by poor comparability in the measurement of socioeconomic status on the death certificate and in the population census. This numerator and denominator bias is discussed more fully in the chapter on evaluating data problems. Because of this data problem, unlinked cross-sectional mortality studies can provide only approximate estimates of the precise magnitude of socioeconomic inequalities in mortality.

The unlinked cross-sectional approach is only feasible if the death registry includes information on the socioeconomic status, most frequently the occupation, of the deceased. In various countries this information is not available, so that it is impossible to categorize mortality by socioeconomic status at the individual level. In these circumstances, it is still nearly always possible to obtain useful indications of the existence of socioeconomic inequalities in mortality by the aggregate-level or ecological approach. In this approach, the study population is divided into smaller or larger geographical units based on place of residence and the association between morbidity or mortality, and socioeconomic status is then studied by comparing the mortality level of these areas with data from other sources on the average socioeconomic status of these areas. The most frequently applied variants of this approach consist of comparing districts within a specific city *(44–49)* and comparing regions within a country *(50–52)*.

A major problem with aggregate-level studies is that it cannot be assumed that mortality differences at the aggregate level (that is between areas with low and high average socioeconomic status) reflect mortality differences at the individual level (that is, between people of different socioeconomic status). This ecological fallacy is described more fully in the chapter on evaluating data problems. The consequence is that ecological studies cannot be used to estimate the size of health differences at the individual level. The value of ecological studies is that they can indicate the effect of socioeconomic disadvantage on health when no individual-level data are available. In addition, comparisons of areas may provide highly relevant information for local health policies, because they identify clearly defined areas with excess health problems. For this reason, ecological studies remain very interesting even in countries or cities where individual-level data are available *(44–46,53,54)*.

Evaluating the informative value of available data

After potential sources of data on socioeconomic inequalities in health have been identified, the next step is to assess the likelihood that these can provide relevant, valid and detailed estimates of the magnitude of inequalities in health, and of changes thereof. The informative value of data sources strongly depends on the ability to measure socioeconomic inequalities in health according to the specifications given in the next chapters. The information contained in these chapters is summarized in Table 3, which can be used to evaluate how informative a mortality study or health interview survey is for the measurement of socioeconomic inequalities in health.

Table 3 consists of a set of questions on the data problems most often encountered in studies of socioeconomic inequalities in health. These questions are ordered in five sections. The first section considers the general characteristics of the data. Sections 2–4 contain questions on the presence of data problems that could lead to biased estimates of the size of socioeconomic inequalities in health at one point in time. Section 5 contains questions that should be considered if trends in health inequalities over time are to be assessed.

Each negative answer to a question in Table 3 means that the informative value of the data source is restricted in some way. The more positive answers, the more informative is the data source.

Table 3. Checklist for the evaluation of the informative value of data sources

1. **Relevance**
   - Does the health interview survey include several indicators of morbidity (at least three)? Do the mortality data allow for a distinction by cause of death?
   - Do the data include at least two of the three essential socioeconomic indicators (occupation, education and income)?
   - Do the data refer to a recent period (less than ten years ago)?

2. **External validity (representativeness)**
   - Are both men and women included?
   - Do the data cover all age groups or at least a very substantial part of the total age range (for example, 15–74 years)?
   - Are you sure that the study is not restricted to a specific region or city or to a part of the working population (such as civil servants or employees of specific companies)?
   - Do the data include all relevant subpopulations (such as people lacking social or health insurance, foreigners or – in case of elderly people – the institutionalized population)?
   - If the study is based on a survey, are you sure that nonresponse does not seriously bias the results?

3. **Internal validity**
   - Can socioeconomic indicators and mortality or morbidity be linked at the individual level?
   - If the mortality study is at the individual level, is it a longitudinal or linked cross-sectional study?
   - If occupation data are available, is occupation measured according to the specifications given in Chapter 3 (Measuring socioeconomic status)? For example, are high and low social classes distinguished with the available data and can social class be determined for most of the economically inactive subjects (such as housewives and retired and disabled people)?
   - If education data are available, is education measured according to the specifications given in Chapter 3? For example, are the lower levels of education (no education, elementary education and lower-secondary education) distinguished?
   - If income data are available, is income measured according to the specifications given in Chapter 3? For example, is the income known for most subjects (at least 80%)?

4. **Precision (statistical power)**
   - Is the health interview survey fairly large (more than 5000 respondents)? Has the mortality study already registered a fairly large number of deaths (more than 1000)?

5. **Relevance and validity of comparisons over time**
   - Do the two points of time that are to be compared cover a sufficiently long span of time (about ten years or more)?
   - Do the health interview surveys that are to be compared have health indicators in common that are based on nearly identical questionnaires? Do the mortality studies that are to be compared have the same basic study design (for example, longitudinal)?
   - Is the measurement of socioeconomic indicators comparable for the two points in time and, in addition, can the same socioeconomic classification be applied to both moments?

In many national and local situations, the evaluation of existing data sources is likely to show that the available data are far from perfect, and that more relevant and valid data are greatly needed. Additional data may be a needed for whenever a question in Table 3 is answered negatively.

Generating new data

At least two procedures can potentially generate substantial new information on health inequalities at a relatively low cost: adding variables to existing data sources and linking data from different registries.

Adding variables to existing data sources. One way is to add health variables to socioeconomic surveys or registries: for example, by including a few questions on health to labour force or household expenditure surveys. Simple, straightforward questions can reveal socioeconomic inequalities in health: for example, "How is your health in general: very good, good, fair, poor or very poor?" and "Do you suffer from any longstanding disease or disability?". The alternative procedure is to add socioeconomic data to health surveys, registries or disease registers. The routine inclusion of socioeconomic data in epidemiological and other health-related research could generate a wealth of reliable and in-depth data on socioeconomic inequalities in health in the long term. The socioeconomic indicator that can be obtained with the least effort and validity problems is education, although in some cases the registration of occupation might be more accurate (perhaps, for example, in proxy interviews and death certificates). A recent experiment in the Netherlands showed that information on both the education and the occupation of patients can be included in routine registries (in that case, hospitals) with fairly little effort and adequate validity *(55)*.

Linking data from different registries. A data registry containing mortality or morbidity data can be linked with a data registry with information on the socioeconomic characteristics of the same population. Successful examples include mortality studies that linked mortality data from vital registries with socioeconomic data from, among others, population censuses *(18–23)*, health interview surveys *(56)*, social security registries *(29,57)* and tax registries *(58)*. A distinction can be made between individual linkage (which identifies individuals by means of a unique identification code) and statistical linkage (which identifies individuals by a combination of sociodemographic characteristics such as date of birth and current place of residence). Especially with individual linkage, provisions may have to be made for complying with confidentiality constraints. With statistical linkage, it is often not possible to distinguish each individual with absolute certainty. For example, a nationwide mortality study in Italy only succeeded in linking 75% of the death certificates with the corresponding records in the population census *(59)*.

It is essential for monitoring purposes that the socioeconomic inequalities in health can be measured repeatedly through time at regular intervals. Efforts should therefore be made to secure the continued availability of data over time. This requires efforts in both organization and methods.

Health surveys and mortality studies need to be organized to ensure that they are repeated at regular time intervals. In addition, future access to data from present studies should be secured.

Comparability in the measurement and classification of socioeconomic and health indicators needs to be ensured across time. Standard socioeconomic classifications should be adopted, and the responsible institutions should be wary of modifying these classifications. Standard socioeconomic classifications have been used for many years in France, Sweden and the United Kingdom. Health interview surveys should continue to use exactly the same interview question for at least some of the health indicators. Even small changes in the wording of questions, their checklists or their response categories can potentially bias comparisons over time *(60,61)*.

Finally, the quest for more valid and more detailed data should not be pursued endlessly. When good descriptive data become available, attention should gradually shift towards explanatory studies and evaluations of health policy interventions. Within Europe, most examples of such studies come from the United Kingdom, the Netherlands and Scandinavia *(24–26,56,62–65)*.

MEASURING MORTALITY AND MORBIDITY

After the sources of information on health inequalities have been chosen, numerous more specific measurement issues should be dealt with. It has to be decided which morbidity and mortality indicators will be used and how the socioeconomic status of subjects will be measured and classified. These measurement issues are discussed in this and the following chapter.

Morbidity

Health interview surveys offer the best prospects for measuring and monitoring inequalities in morbidity in most countries and regions of Europe *(66)*. Such issues as optimal sample size and sampling techniques have to be resolved, but since these issues are covered in specialized books, they are not discussed here.

Health interview surveys can include a broad spectrum of health indicators related to general health status and problems with physical health, mental health or social functioning *(11,12)*. The choice of the indicators to be included depends largely on such considerations as the relevance to the local situation and feasibility. For example, if there is only space for a few questions, the choice is restricted to indicators based on one or a few questions (such as "How is your health in general?"), whereas indicators based on a battery of questions have to be dismissed. The possibilities for monitoring inequalities in health would be greatly improved if new health surveys included an identical version of one or more of the health indicators included in surveys conducted previously.

An example of a study that used a balanced series of health indicators is the Health and Lifestyle Study from England and Wales *(67)*. In this study, four components of health were defined and measured:

- experience of illness or freedom from illness, which was measured by means of a checklist of 16 symptoms, such as headaches, colds and flu, and painful joints;
- psychosocial malaise or wellbeing, which was measured with a checklist of eight symptoms, such as nerves, always feeling tired, and feeling lonely;
- disease or disability or their absence, which was measured with a series of questions that screened individuals as to whether they had a nonlimiting or limiting disability and then established the severity of any handicap they experienced; and
- fitness, a term used in the Health and Lifestyle Study to summarize three physical measures of health registered by a nurse: blood pressure, body mass index ((weight in kg/height in metres)^2) and respiratory function. In addition, an overall index of health status was constructed based on these four components of health.

An international network under the aegis of WHO has recommended a number of health indicators that have shown to be applicable in a large number of countries and that could serve as indicators for monitoring progress towards several European targets for health for all *(68)*.

Widespread application of these indicators, which are given in Table 4, would greatly contribute to the international comparability of data on morbidity in general and on inequalities in morbidity in particular.

Table 4. Instruments for health interview surveys recommended for monitoring European health for all indicators

Health for all indicator (number)	Health indicator (general content)
Perceived health (2.2)	Percentage of respondents saying that their general state of health is very good, good, fair, poor or very poor
Temporary disability (4.1)	Average number of days that respondents had to cut down their usual activities because of: • physical disability • mental conditions
Long-term disability (4.2)	A summary score derived from questions on: • confinement to bed, chair or house • difficulty with walking, dressing, washing, feeding, hearing, seeing, etc.
Prevalence of selected chronic conditions (4.6)	Physical health: no recommended instruments are yet available Mental health: batteries of questions on, respectively, dementia, mental retardation and mental disorders
Emotional wellbeing (recommended new indicator)	General Health Questionnaire: a summary score derived from questions on worry, concentration, feeling happy, self-confidence, etc.

Source: adapted from *Third consultation to develop common methods and instruments for health interview surveys:* report on a WHO meeting *(68)*.

Morbidity rates are expressed as the number of health problems in relation to the number of respondents. This relationship is usually expressed as a proportion (the proportion of respondents that report a health problem) or as a count (the number of reported health problems per 100 respondents). Count variables can also be expressed as a proportion but at the expense of loss of information. For example, the prevalence of chronic conditions, which is expressed as the number of conditions per 100 people, can also be expressed as the percentage of respondents reporting one or more conditions, but that figure does not take into account the exact number of conditions reported by those who have one or more chronic conditions.

When morbidity rates are being calculated according to socioeconomic group, the effect of any confounding variable should be removed, for example, by standardization *(69)*. Age and sex are obvious confounders, as both are likely to be associated with socioeconomic status and with morbidity. For example, since older people on average have lower levels of education than younger people and old age is also associated with increased disability, the failure to control for age would give a biased (too high) estimate of the association between low education and disability. Age and sex standardization is therefore generally required before any meaningful comparison of morbidity rates between socioeconomic groups is possible. Under some circumstances, other variables may also need to be controlled for, such as ethnicity and rural versus urban residence.

Statistical techniques such as standardization must not be used thoughtlessly, however, because they may conceal important variation in the size of health inequalities according to age, sex, place of residence and other parameters. Sometimes it is better to present separate estimates for subgroups of the population than to pool all differences in one standardized figure. For that reason, it is common to distinguish men and women in all measurements of health inequalities.

Mortality

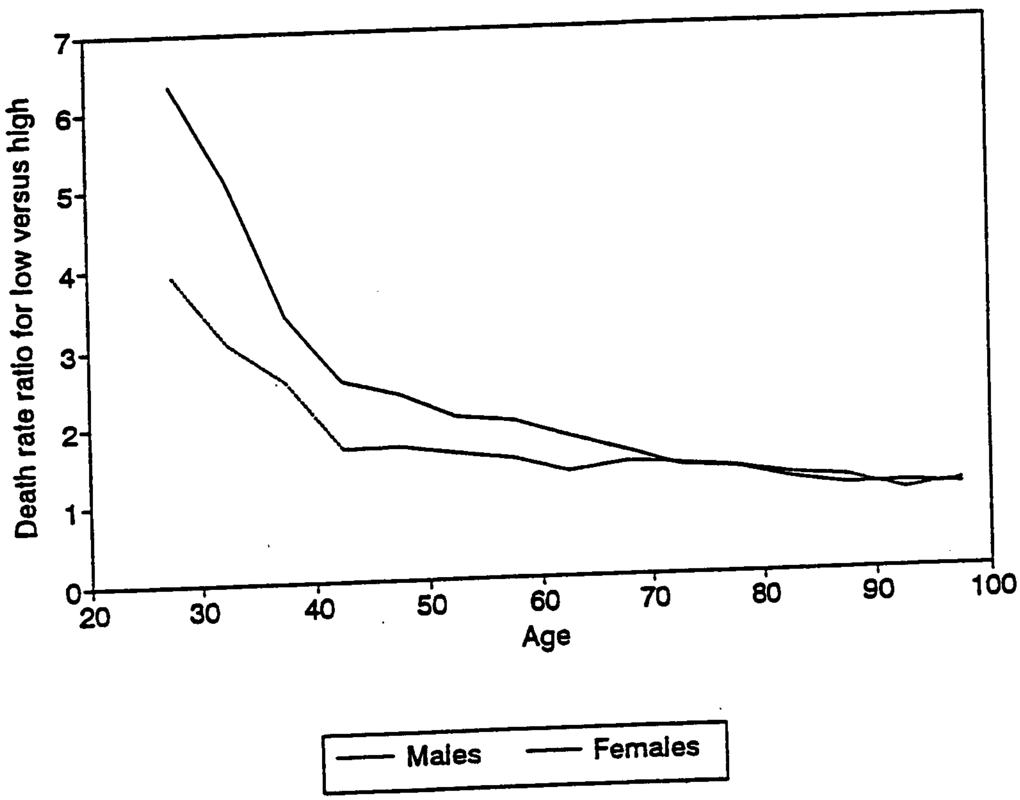
Mortality levels are basically expressed as incidence rates, which are calculated by dividing the observed number of deaths by the corresponding number of person-years (the number of people times the average number of years per person) exposed to the risk of dying *(15–17)*. In longitudinal studies, the number of person-years can be calculated accurately from the available data. In unlinked cross-sectional mortality studies, it is customary to estimate this number as the number of people in the middle year of the study period times the number of years covered by the study period.

As with morbidity, the effect of confounding variables should be removed by means of standardization or other techniques. Age and sex are the most obvious variables, but in certain circumstances other variables may also need to be controlled for. For example, ecological studies on differences in mortality between affluent and nonaffluent urban districts should correct for the fact that most cities have a few districts with large institutions (such as homes for elderly people) to which ill and disabled people move, which makes the death rates in these districts exceptionally high *(45)*.

As with morbidity, statistical techniques such as age standardization should not be used thoughtlessly, and sometimes it is better to present separate estimates for subgroups of the population. Since mortality differences often vary strongly by age and sex, it is usual to present inequality estimates by sex and by broad age group (for example, 0–14, 15–44, 45–64 and 65 and over years).

Finland provides an example of variation in the size of health inequalities by sex and age. Finland is one of the few countries with valid data on inequalities in mortality across the entire age range *(18)*. The differences in mortality (at least in relative terms) in Finland are largest for young age groups and gradually diminish with increasing age (Fig. 2). Before the age of 70 years, inequalities in mortality are about twice as large for men as for women. In other countries, inequalities in mortality are larger for men than for women and change considerably with increasing age *(70)*.

Fig. 2 The mortality difference (expresses as a ratio) between people with elementary education only and people with postsecondary education in Finland, 1986–1990: variation by age and sex.



Source: Valkonen et al. *(18)*.

Study of specific age groups is also required if one wants to distinguish premature deaths from deaths in old age, which cause fewer years of life lost. This line of thought can further be expanded by calculating comprehensive measures such as the life expectancy (at birth or at any other age) or the potential years of life lost *(18,38,71,72)*. A major advantage of these measures is their straightforward interpretation *(73)*. Nevertheless, such measures cannot be calculated when data are only available for a restricted age range. A good alternative, then, is to express mortality levels in terms of the probabilities of death and the chance of survival. For example, if the study is restricted to the age group 35–60 years, the mortality level could be expressed as the chance that a person aged 35 years will die before becoming 60 years old. This was how the main results of a national longitudinal study in France were summarized for the period 1976–1980: only 1 of 14 professionals but almost 1 of 5 unskilled labourers died between age 35 and 60 years *(23)*.

A special case is mortality during the perinatal period and in infancy. The experience of various countries, including Belgium, Israel and Luxembourg, indicates that inequalities in infant and perinatal mortality can be measured and monitored even when this is not possible for mortality at other ages *(74–76)*. Therefore, when no data are available on inequalities in adult mortality, it may be worthwhile to make special efforts to obtain data on perinatal and infant mortality. Although mortality at these ages cannot be regarded as a comprehensive measure of ill health, the relevance of perinatal and infant mortality is that they have shown to be strongly sensitive to socioeconomic disadvantage *(77)* in a wide variety of situations.

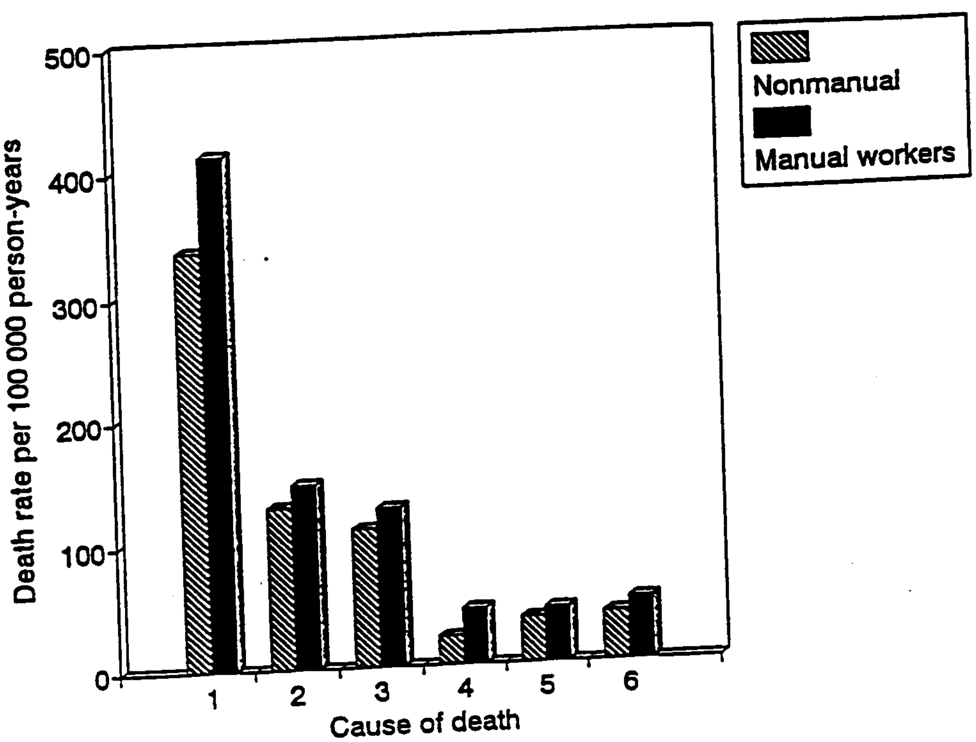
Another line of research on mortality deals with specific causes of death. The study of specific causes has two purposes: it is often the only way to obtain indications of the magnitude of the inequalities in the prevalence or incidence of specific diseases, and study of such specific causes of death as lung cancer, liver cirrhosis and accidents provides clues to the explanation of inequalities in all-cause mortality. A few general guidelines can be given on the most appropriate way of classifying causes of death. If the distinguished cause-of-death groups are broad and heterogeneous (such as all neoplasms or all cardiovascular diseases), they say very little about the causes of health inequalities. A very detailed classification (such as the three-digit codes of the International Classification of Diseases (ICD)) produces an overwhelming quantity of data and easily runs into problems related to small numbers of deaths. A convenient classification is obtained by distinguishing the 17 ICD chapters and by further distinguishing within the largest ICD chapters, for example the ten largest single causes of death.

Fig. 3 gives an example of the informative value of distinguishing causes of death. In a longitudinal study in Turin, Italy, the size of social class differences was estimated for six broad groups of causes of death *(28)*. Within each group, the death rates were higher for manual workers than for nonmanual workers. The largest relative difference, a twofold difference, was observed for deaths from diseases of the digestive system, which are mainly caused by liver cirrhosis. Large inequalities in mortality from liver cirrhosis were also observed in a nationwide Italian study *(59)*, suggesting that alcohol abuse contributes substantially to inequalities in mortality in Italy.

Combining morbidity and mortality

Such measures as disability-free life expectancy and healthy life expectancy *(78)* suddenly became more popular in the 1980s. Disability-free life expectancy has even been included in the list of health for all indicators *(6)*. These measures are attractive because they combine information on morbidity and mortality into one single figure with a clear and attractive interpretation. For example, if this measure is calculated using data on disability and for all age groups, it represents the average number of years without disability that people may be expected to enjoy during their entire life, assuming that they would be exposed during their entire lifetime to the age-specific morbidity and mortality rates that are observed now.

Fig. 3 Differences in cause-specific mortality for selected causes of death between manual and nonmanual employed men 15–59 years old in Turin, Italy, in 1981–1985



1 = all causes
2 = neoplasms
3 = cardiovascular diseases
4 = diseases of the digestive system
5 = all other diseases
6 = external causes

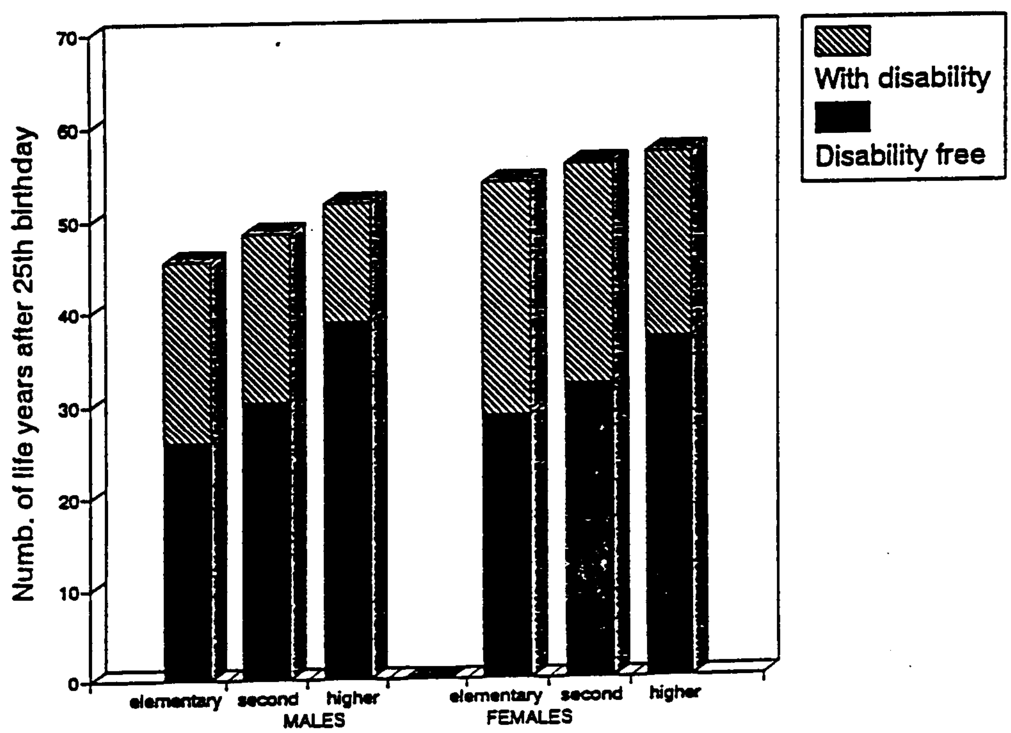
Source: calculated from data presented in Costa & Segnan *(28)*.

The measurement of disability-free life expectancy is still being refined, and new measures such as disability-adjusted life years may obtain wide acceptance and replace disability-free life expectancy *(79)*. Because consensus on various methodological issues has not yet been reached, the further discussion is limited to the disability-free life expectancy, which is most frequently calculated by the Sullivan method.

Calculating the disability-free life expectancy for specific socioeconomic groups is a promising way to summarize the available information on health inequalities, but the disability-free life expectancy still lacks widespread use in the field of health inequalities. A reason for this is that the calculation of disability-free life expectancy for specific socioeconomic groups requires that high-quality data be available on both mortality and morbidity, that the same socioeconomic classification be used in both types of data and that the data refer to a broad age range, preferably all age groups.

One of the countries in which socioeconomic variation in the disability-free life expectancy can be estimated is Finland *(80)*, where a low educational level is associated with a shorter life expectancy after age 25 years (Fig. 4 total bar). People with lower educational levels not only die younger but suffer more years of longstanding and disabling illness (Fig. 4 black bar). As a result, these people enjoy fewer years of good health (Fig. 4 striped bar). The difference between the lowest and highest level of education is more than 8 years of disability-free life expectancy for women and 13 years for men.

Fig. 4 Life expectancy with and without disability after age 25 years according to sex and educational level in Finland, 1986–1990



Source: Valkonen et al. (80).

MEASURING SOCIOECONOMIC STATUS

Within each society, material and other resources are unequally distributed. This inequality can be portrayed as a system of social stratification or social hierarchy *(8)*. People attain different positions in the social hierarchy according to their occupational status, educational achievement and income level. Their position in the social stratification system can be summarized as their socioeconomic status.

Although occupation, educational level and income together determine a person's position in the social hierarchy, these factors are sufficiently distinct to require that they be studied separately in relation to health. Educational level creates differences between people in terms of access to information and the level of proficiency in benefiting from new knowledge, whereas income creates differences in access to scarce material goods. Occupational status includes both these aspects and adds to them benefits accruing from the exercise of specific jobs, such as prestige, privileges, power and social and technical skills.

This chapter provides general guidelines for the measurement and classification of occupation, education and income to assess socioeconomic inequalities in health. For details on the measurement of occupation, education and income, including the wording of questions to be included in surveys, see the specialized literature *(81,82)*. This chapter ends with a discussion of the use of proxy measures for socioeconomic status such as housing tenure and car ownership.

Occupational status

Occupation is relevant because it determines people's place in the social hierarchy and because it indicates exposure to specific occupational risks, such as toxic compounds. The main issue then is how to classify people with specific jobs according to their place in the social hierarchy. This section discusses the two main approaches and briefly discusses specific issues related to the classification of women, children and economically inactive men.

Classifying people by social class

The most usual approach consists of classifying people based on their position in the labour market into a number of discrete groups or social classes. In its basic form, this approach distinguishes:

- higher-level employees, employers and professionals
- lower-level employees
- skilled labourers
- unskilled labourers
- farmers
- other self-employed workers.

People can be assigned to social classes by means of a set of detailed rules that use information on such items as occupational title, skills required, income pay-off and leadership functions.

Many countries have their own variant of a social class scheme. One rather well known scheme is the Registrar General's scheme of social classes, which has been used in England and Wales since the beginning of the twentieth century: (I) professional, (II) intermediate, (III NM) skilled nonmanual, (III M) skilled manual, (IV) semiskilled manual and (V) unskilled manual *(83)*. Although the Registrar General's scheme has been criticized for a lack of theoretical foundation, its application was exemplary in at least one sense: this class scheme has been consistently used across many fields of research and over a long period of time. This has permitted a number of well known studies of trends in inequalities in mortality in England and Wales *(34–38)*.

To overcome differences in classifications between countries, which hamper international comparative studies of social stratification and social mobility (and socioeconomic inequalities in health), Erikson, Goldthorpe and Portocarero have developed an elaborated social class scheme: the EGP scheme *(84)*. This scheme is illustrated in Table 5 for the three countries to which it was first applied (England, France and Sweden). An approximation to this scheme can be applied whenever information is available on occupational title (bricklayer, accountant, etc.), employment status (self-employed or employed) and supervisory status (the number of subordinates) *(85)*. The EGP scheme has been applied in various countries and is considered to be able to distinguish groups that differ in various respects, such as living conditions and social mobility.

Not all social classes have a clearly hierarchical relationship to each other *(84)*. For example, no general rule says that lower-level employees have a higher position in the social stratification system than skilled labourers; although lower-level employees may have a greater affinity or proximity to higher employees than skilled labourers *(85–87)*, they differ little from skilled labourers in income gained, educational requirements and occupational prestige. Similarly, the position of farmers depends much on the local situation and, in general, is not clearly higher or lower than that of other self-employed workers. Therefore, not all health differences between social classes are equally relevant to the study of socioeconomic inequalities in health, and of most interest are the health differences between classes with a clearly different position in the social hierarchy. Examples of the latter are health differences between higher and lower employees, between skilled and unskilled labourers and between self-employed people with and without subordinates. The most comprehensive comparison is often that between middle and high-level employees versus semiskilled and unskilled workers (including farm labourers).

Table 5. The EGP social class scheme illustrated for England, France and Sweden

	England	France	Sweden
I	Higher-grade professionals, administrators and officials; managers in large industrial establishments; large proprietors	Professions libérales, littéraires et scientifiques; professeurs; ingénieurs; cadres administratifs supérieurs; industriels; gros commerçants	Industriidkare, högre förvaltningspersonal i enskild och offentlig tjänst, fria yrken inom akademiska professioner
II	Lower-grade professionals, administrators and officials; higher-grade technicians; managers in small business and industrial establishments; supervisors of nonmanual employees	Instituteurs et professions intellectuelles diverses; services médicaux et sociaux; techniciens; cadres administratifs moyens; artistes; clergé et armée; police	Tjänstemän på "mellannivå" i offentlig och enskild tjänst, folk- och grundskolelärare, högre huslig tjänst, fartygsbefäl
III	Routine nonmanual employees in administration and commerce; sales personnel; other rank-and-file service workers	Employés de bureau; employés de commerce; gens de maison; autres personnels de service	Kontorspersonal, vissa handelsbiträden, servicepersonal
IVa	Small proprietors; artisans, etc. with employees	Petits commerçants, artisans avec employés	Småföretagare, hantverkare, handelsmän och skeppare med anställda
IVb	Small proprietors; artisans, etc. without employees	Petits commerçants, artisans sans employés	Småföretagare, hantverkare, handelsmän och skeppare med anställda
IVc	Farmers and smallholders; self-employed fishermen	Agriculteurs exploitants; patrons pêcheurs	Godsägare, hemmansägare, småjordbrukare, arrendatorer, självständiga fiskare
V/VI	Lower-grade technicians; supervisors of manual workers; skilled manual workers	Contremaîtres, ouvriers qualifiés; mineurs; marins; pêcheurs	Verkmästare och förmän för arbetare, sjömän och arbetare i offentlig och privat tjänst i arbeten som kräver yrkeskunskap
VIIa	Semi- and unskilled manual workers (not in agriculture)	Ouvriers spécialisés; manœuvres	Sjömän och arbetare i offentlig och privat tjänst i arbeten som ej kräver yrkeskunskap, lägre huslig tjänst, vissa affärsbiträden
VIIb	Agricultural workers	Salariés agricoles	Jordbruksarbetare, skogsarbetare, anställda fiskare

Source: Erikson & Goldthorpe *(84)*.

Classifying occupations by means of one-dimensional scales

The above-mentioned complexities in the study of health variations between social classes could be avoided to some extent by using strictly hierarchical classifications, which unambiguously rank occupations along one single scale. Such classifications can only be used if one assumes that the most relevant differences between occupations can be captured in one dimension *(88)*. An attractive feature of this approach is that it results in the construction of a continuous scale, which means that it allows for an unlimited number of graded distinctions between occupations, including between occupations that otherwise would be combined into the same social class. Two types of scales are common in social stratification research.

- Prestige scales express the general desirability of occupations and are derived from the subjective judgement of lay or expert panels. This measure stresses the reward dimension of occupation, similar to and sometimes compensating for income.

- Scales based on income and education are more readily relevant to health. These scales are constructed as a weighted sum of the average education and average income of people with a certain occupation. They intend to measure the potential of an occupation to convert a person's main resource (education) into a person's main reward (income).

Internationally applicable versions of both types of scales have been developed *(88,89)*, whereas national scales are available for various countries. These scales are often used in social stratification research, but their relevance to health has not yet been assessed. A study in Israel found that the association between occupational status and morbidity is equally strong when assessed either with the social class approach or with a one-dimensional scale *(90)*.

Classifying children, women and economically inactive men

Economically inactive men (unemployed, disabled, retired and homemakers) are often studied separately by putting them in one or a few inactive groups. This is often the only thing that can be done, because additional socioeconomic information on economically inactive men is lacking. Nevertheless, it is more appropriate to the study of health differences across the social hierarchy to estimate the place in the social hierarchy of all economically inactive people individually. Based on the assumption that the social position of most economically inactive men is largely determined by their last or longest held occupation, information on previously held occupations, if available, can be used to assign economically inactive men to the corresponding social class or hierarchical position *(91)*.

Children are classified according to the occupation of their parents, either their mother or their father *(92)*. The importance of either parent probably varies according to country or region and to the health problem being analysed.

For health differences among women, the fundamental question is whether the occupational status of women should be derived from their own occupation or whether the partner's occupation is a better reference point for married or cohabiting women that are homemakers or that work part-time or in a low-status job (this dilemma also applies to some men). In theory, both perspectives are valid and there is no clear preference for either one *(93,94)*. In practice, the choice between the two methods largely depends on national customs and the availability of data. Information on a woman's own occupation is usually more readily available than information on the occupation of the partner. If all women are classified by their own occupation, however, several problems have to be resolved. First, lack of information on previous occupations may make it impossible to assign an occupational status to homemakers. Second, the social class analysis may need to be adjusted for men when applied to women, as the occupational distribution of women is different *(95)*. For example, class analyses need to take into account the fact that the majority of women in some social classes often work in a few typically female occupations such as nurses, teachers and secretaries.

Smaller health inequalities are often observed when all women are classified according to their own occupations than when cohabiting and married women are classified according to their partner's occupation. An example from Great Britain is presented in Fig. 5 *(93)*. Application of the traditional approach, classifying married women according to their husband's occupation, and single women to their own occupation, demonstrates large and consistent differences in morbidity between social classes. When all women are classified according to their own occupation, morbidity differences are smaller and less consistent: the difference between skilled and unskilled labourers is small, and there is no difference between high and low professionals.

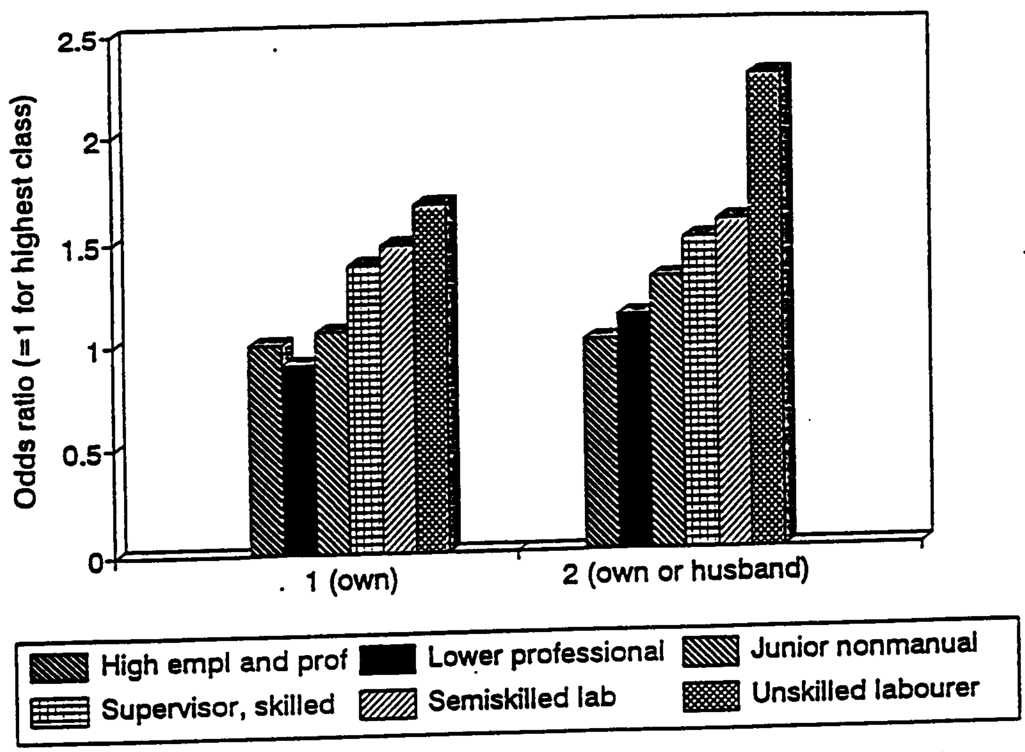
Educational level

As with occupation, information on education is used to distinguish people with a high position in the social hierarchy from those with a low position. The relevant aspect of education is therefore the level of education. The most direct measure is the highest level of education that has been successfully completed (with diploma, if applicable). If possible, this measure takes into account not only general education but also technical and vocational education, and not only full-time education but also part-time study after leaving school.

The precise classification of levels of education depends on the system under which the subjects have received their education. This classification is usually sufficiently detailed if the following categories are distinguished:

- no education completed
- primary (elementary) level
- secondary level, first (lower) stage
- secondary level, second (upper) stage
- third (postsecondary) level.

Fig. 5 Social class differences in the prevalence of limiting longstanding illness in Great Britain among women 16–64 years old, 1985–1987. Comparison of two approaches of classifying women



1 = all women classified according to their own occupation
2 = married women classified according to husband's occupation

Source: calculated from data presented in Arber & Ginn *(93)*.

In most educational systems, the formal ages at which the last four educational levels are completed are approximately 12, 15, 18 and 21 or more years respectively *(96)*. Table 6 illustrates the application of this basic scheme to the educational system of three countries.

When no information is available on the level of education completed, substitute measures include the highest level of education that a person has attended but not necessarily completed or the number of years that a person attended school (an approach that is often used in the United States *(97,98)*). The latter figure has the attractive property of being a quantitative measure of socioeconomic status, but in its most simple form it fails to take into account the type, and therefore the level, of education that was attended.

Table 6. Examples of classifications of four broad educational levels: Federal Republic of Germany, Greece and Spain

Broad educational level	Federal Republic of Germany	Greece	Spain
Elementary level	Grundschulen (4 years)	Dimotiko (6 years)	Educación general básica, ciclo medio (5 years)
Secondary level, first stage	Realschuler, Hauptschulen (5–6 years)	Gymnasio (3 years)	Educación general básica, ciclo superior (3 years)
Secondary level, second stage	Gymnasien, Fachoberschuler, Berufsschulen (3 years)	Geniko, Lykeo Techniko-Epagelmatiko (2–3 years)	Bachillerato, Formación profesional grado 1 o 2, Orientación universitaria (3–5 years)
Third level	Universitäten, Hochschulen (up to 5 years)	Anotato Ekpedeutiko Idryma, Epagelmatiko Idryma (up to 7 years)	Enseñanzas universitarias, Escuelas técnicas superiores (up to 7 years)

Source: Education in OECD countries 1987–89. A compendium of statistical information (96).

Income level

The income level of a person can be used in two ways. Income indicates the socioeconomic status of the income recipient, with higher personal income indicating, among other things, a better position in the labour market. Income is more commonly used complementarily to occupation and education, however, and as a more proximate indicator of access to scarce material resources or of standard of living.

The standard of living can be expressed most adequately when the income level is measured by:

- adding all income components (this yields total gross income)
- subtracting deductions of tax and social contributions (net income)
- adding the net incomes of all household members (household income)
- adjusting for the size of the household (household equivalent income).

Many methods have been developed to take into account the size and – less often – the age composition of households. For various countries, standard formulae have been developed, and it is often most appropriate to follow these national rules. A simple formula that has been used in an international study on poverty and income inequality consists of dividing the household income by the number of household members (n) to the power 0.36 ($n^{0.36}$) *(99)*.

A practical problem in the measurement of income in interview surveys is the high nonresponse rate to questions about personal and household income. Moreover, willing respondents may not report income accurately or validly. Measuring income accurately requires a large battery of questions, but most health interview and similar surveys have no room for an extensive measurement of income. In these cases, the survey should at least attempt to determine the income of each adult in the household and to make sure that the respondents take into account the most relevant income components such as wages and salaries, interest, pensions and transfer payments.

The method of classifying income levels can follow national customs. Income classifications should be detailed enough so that the frequency of health problems in the most materially deprived groups can be studied. A convenient way to achieve comparability over time is to correct income levels for inflation (yielding real income) and then to apply the same classification of real income to both points in time.

Proxy measures for socioeconomic status

The main value of proxy measures is that they can indicate the existence of health inequalities when data on health differences by occupation, education or income are lacking. A wide variety of proxy measures can fulfil this purpose. The two most important types of measures are discussed in this section.

One often used category of proxy measures relates to material living standards. Two indicators that have gained wide popularity within the United Kingdom are car ownership and housing tenure *(100)*. More widely used indicators include the quality of housing, such as crowding (number of people per room) or the absence of basic amenities (bath, toilet and running drinking-water) *(101,102)*. In some countries, entitlement to social security or to health insurance depends on people's position in the labour market and income levels, and this entitlement can therefore also serve as a proxy measure for socioeconomic status *(103,104)*. These indicators have some advantages over income as a measure of living standards. First, they are more stable over time; income may vary substantially, even in the short run, due to changes in household composition or in the employment and remuneration of household members. Second, questions on house ownership, car access, entitlement to health insurance, etc. do not create the problems of nonresponse and inaccuracy that are typical for survey questions on income.

Nevertheless, the validity of these measures as indicators of socioeconomic status is likely to be limited and depends on the strength of their association with occupation, education and income. Even if one interprets these proxy measures in a narrower sense as indicators of material living standards only, without reference to the broader concept of socioeconomic status, the crudeness of these measures probably limits their validity. The association between these variables and morbidity or mortality may reflect the effect of low living standards and other health-related factors as well. For example, the often reported lower mortality among homeowners probably reflects not only the higher living standards of homeowners but also the differences between them and tenants in geographical mobility and the motivation to invest in the future as well as selection for health in the availability of mortgages.

Another category of proxy measures consists of sociodemographic variables that define sections of the population that live in a special situation of disadvantage: unemployed people, ethnic minorities, single mothers and people who depend on social assistance benefits *(105–107)*. The choice of the groups of interest depends on the national context. For example, several studies from Israel focus on health differences associated with country of birth *(108)*. If socioeconomically disadvantaged groups are found to suffer from high morbidity and mortality rates, this indicates that socioeconomic disadvantage is related to bad health. Whether such evidence can be used as a substitute for studying the broader inequalities in health across the entire social hierarchy depends on the same validity considerations as those given above.

Fig. 6 illustrates the problems of applying proxy measures using the data from a study on infant mortality in Greece *(109)*. Large mortality differences are revealed when the educational level of either the mother or the father is used as the socioeconomic indicator. In contrast, proxy measures were not able to demonstrate large mortality differences: whether or not the mother had a health insurance did not make any difference, and the mortality differences associated with the quality of housing were much smaller than those associated with parental education. In this example, therefore, sole reliance on proxy measures would have given the erroneous impression that socioeconomic status is not strongly associated with the risk of perinatal death.

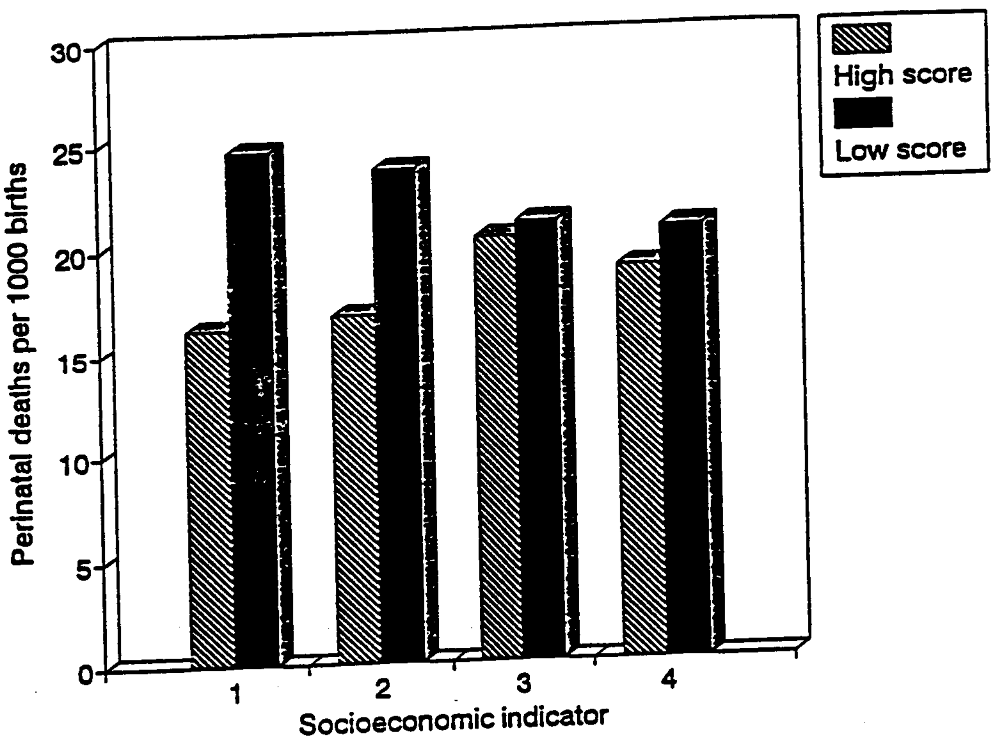
Proxy measures may roughly indicate the existence of health inequalities but cannot measure magnitude of socioeconomic inequalities in health. It is also important that proxy measures considered to be reasonably valid indicators of socioeconomic status in one situation may be invalid in other situations.

The measurement of socioeconomic status in ecological studies

Ecological studies represent a special case in the measurement of socioeconomic status. In ecological studies, the morbidity and mortality rates of a region or an urban district are related to the average socioeconomic status of the residents in these areas, and the observed associations are used, among others, to indicate the existence of an association at the individual level *(44–49,53,54)*. The central question is, then, how to measure the socioeconomic level of the residents of an area.

Occupation, education and income are often closely correlated at the area level. In addition, a low average socioeconomic level is often reflected in a host of other indicators, such as poor

Fig. 6 Differences in perinatal mortality in Greece in 1983 according to parental education and two proxy measures for socioeconomic status



1 = education of mother (≤ 6 versus ≥ 7 years)
2 = education of father (≤ 6 versus ≥ 7 years)
3 = whether or not mother has health insurance
4 = quality of housing (people per room) (≤ 1.5 versus > 1.5 people per room)

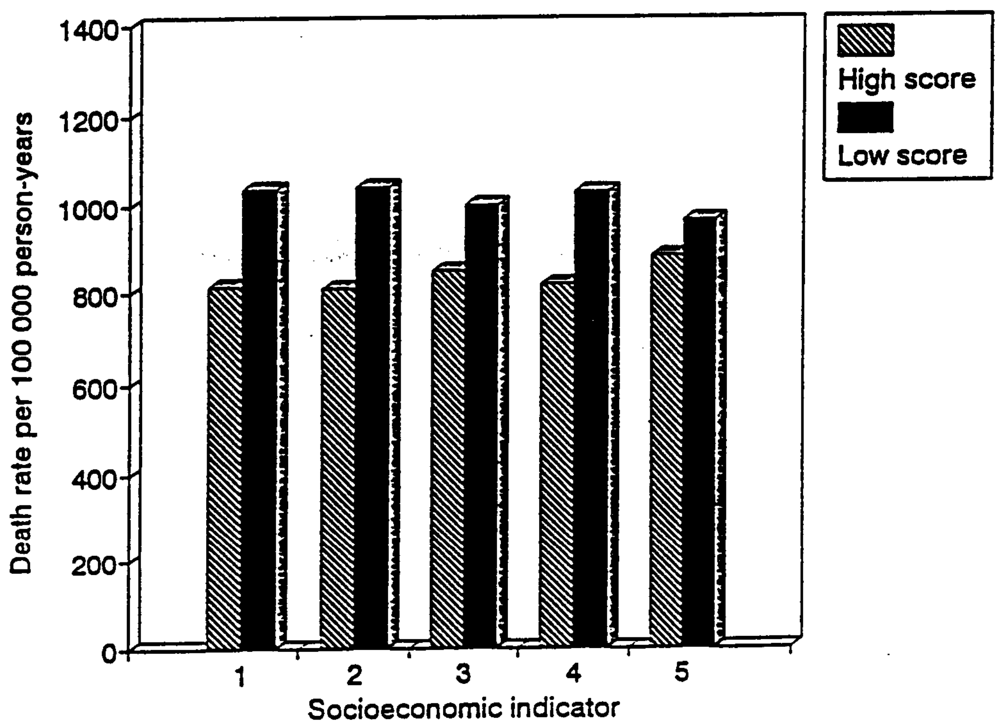
Source: Goldblatt & Fox *(100)*.

housing conditions, high unemployment rates and higher proportions of immigrants or ethnic minorities. One implication is that it is often preferable to use composite measures that combine data on occupation, education and income into one figure. In addition, if data on one or more of these three indicators are lacking, it is often valid to include data on proxy measures in the construction of the composite socioeconomic index.

The United Kingdom has ample experience with the development and application of composite measures of the socioeconomic level of areas, in the framework of ecological studies of health inequalities. Important examples of such composite measures are the Townsend index *(44)* and the Carstairs index *(45)*, which are both considered to be measures of deprivation and take into account area characteristics such as the unemployment rate and the proportion of the population in the lowest occupational class. Both indices have been shown to correlate quite strongly with, for example mortality *(110)*. Similar approaches to the measurement of the socioeconomic level of areas have been applied in other countries *(47)*.

Fig. 7 illustrates the use of various measures of the socioeconomic level of areas in Barcelona *(48)*. About equally large mortality differences between deprived and affluent districts have been demonstrated by using income, education or occupational status as the indicator of the average socioeconomic level of districts. One proxy measure, the unemployment rate, was equally able to reveal the large mortality difference between deprived and affluent districts. Nevertheless, the differences were much smaller when a composite indicator of housing quality was used, even though that indicator combined various aspects of housing quality, such as crowding and basic amenities. This example shows that the use of proxy measures in ecological studies may sometimes, but not necessarily always, produce valid results.

Fig. 7 Differences in mortality per 100 000 person-years between districts within Barcelona for all ages and both sexes in 1984–1987 using five different indicators of the average socioeconomic level of districts



1 = the percentage of people with net annual income greater than Ptas 1 million (approximately US $12 000 using current exchange rates and accounting for inflation)
2 = the percentage of people with elementary education completed
3 = the percentage of people in nonmanual job
4 = the percentage of economically active and unemployed
5 = the percentage of homes with high housing standards

Source: calculated from data presented in Mompart & Pennia *(48)*.

MEASURING THE ASSOCIATION BETWEEN SOCIOECONOMIC STATUS AND MORBIDITY AND MORTALITY

When data on socioeconomic inequalities in health are available, either from existing data sources or as the result of a new data collection effort, the magnitude of these inequalities can be measured. This chapter presents guidelines on the quantitative techniques that can be used for this purpose. This is done in two parts: the first part focuses on the inequalities as they are observed at one point in time, and the second part focuses on changes over time in the inequalities. In both cases, the recommended approach starts by describing in detail the patterns of association between socioeconomic status and morbidity or mortality and, if appropriate, by calculating one or more summary indices.

Describing health inequalities at one point in time

A detailed description of the association between socioeconomic status and health starts by calculating the morbidity and mortality levels for each socioeconomic group. The classification of socioeconomic groups should contain a hierarchical component: that is, it should be possible to distinguish people with a low position in the social hierarchy from those with a high position. In addition, the classification should be detailed enough to allow specific groups with excessive morbidity and mortality levels to be identified.

Box 1 illustrates the recommended approach for describing variation in morbidity and mortality rates by socioeconomic group. This inspection of variation between groups is greatly facilitated by graphical representation. Variation between groups can numerically be expressed as rate ratios or rate differences. Rate ratios express the morbidity or mortality rate of a socioeconomic group as the ratio of the rate of some reference group (set at 1 or 100). Rate differences express the absolute difference in morbidity or mortality between a socioeconomic group and the reference group. For both rate ratios and rate differences, an appropriate reference group is the highest socioeconomic group, if that group is not too small. Sometimes a sufficiently large reference group can only be obtained by combining the highest with the next highest socioeconomic group.

A key element in the basic description is the pattern of the association between morbidity and mortality rates and socioeconomic status. Do rates increase or decrease consistently with decreasing socioeconomic status? Sometimes mortality and morbidity rates do not increase regularly: for example, when specific groups have higher rates than could be expected from their position on the social hierarchy or when the association between socioeconomic status and morbidity or mortality is J-shaped. An example of a nonlinear association relates to income: in a health survey in England and Wales, increasing income was associated with a decrease in morbidity rates but, above a certain income level, morbidity rates remained at the same level *(67)*.

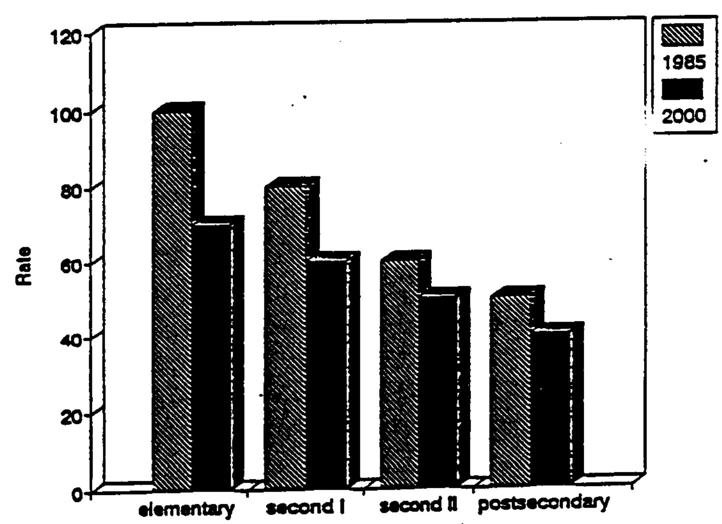
BOX 1 Basic description of health inequalities at one point in time

The table below contains a fictitious example of socioeconomic inequalities in health at one point in time. It describes differences in the prevalence of health problems by educational level in a population of 100 000 people. Column 4 indicates the prevalence rate in each of the four educational groups calculated from the numbers of people with health problems (column 3) and the total number of people (column 1). To reduce the complexity in this example, it is assumed that the age composition of each of the four groups is identical, so that age standardization is not necessary. Both rate ratios (a measure of relative differences, column 5) and rate differences (a measure of absolute differences, column 6) have been calculated from the figures in column 4.

Educational level	Number of people	Population share	Number of people with health problems	Rate of health problems (per 10 000 population)	Rate ratio	Rate difference
	(1)	(2)	(3)	(4)	(5)	(6)
Postsecondary	10 000	0.10	50	50	1.0	0
Secondary, upper level	20 000	0.20	120	60	1.2	10
Secondary, lower level	30 000	0.30	240	80	1.6	30
Elementary	40 000	0.40	400	100	2.0	50
Total	100 000	1.00	810	81	NA	NA

NA = not applicable

The rate of health problems among individuals with the lowest educational level is twice as high as that among those with postsecondary educationwhich was chosen as the reference category. In absolute terms the difference between these two extreme groups is 50 cases of health problems per 10 000 population. The rate ratios and rate differences of the intermediate categories are in between. The rate increases consistently with decreasing educational level.



Summarizing the magnitude of health inequalities at one point in time

Introduction

As a next step to the basic description of health inequalities, summary indices may be calculated that express the size of health inequalities in one single figure. Such figures not only facilitate comparison, such as over time, but may provide a means to translate the magnitude of inequalities into concrete terms that are readily grasped by non-statisticians and to take into account the population size of groups with high morbidity and mortality rates.

Summary indices should always complement instead of replace the basic description of health inequalities. The estimates for summary indices should be checked against the basic data, and any apparent discrepancy between the two types of information should be a reason to determine whether the summary index adequately represents the observed variation between socioeconomic groups in morbidity or mortality.

An assumption underlying the use of all summary indices is that the morbidity and mortality rates of groups are systematically related to their position in the social hierarchy, since summary indices intend to quantify this systematic association. The systematic nature of an association can take many forms: with increasing socioeconomic status, morbidity and mortality rates can decrease (or increase) according to a linear, curvilinear or more complex, such as J-shaped, pattern. If no systematic pattern emerges from the basic data, there is no point in summarizing.

Some summary indices go even further in their assumptions on the association between socioeconomic status and morbidity and mortality. Examples are regression-based indices, which assume that (transformed) morbidity and mortality rates decrease (or increase) linearly with increasing socioeconomic status. Substantial nonlinearity could be a reason *not* to apply regression-based indices.

Summary measures: a threefold classification

Many summary measures for the magnitude of socioeconomic inequalities in health have been applied in the international literature. Most applications aimed at some form of comparison, either over time or between national or subnational populations. A systematic analysis of the properties of these measures reveals that they differ substantially in a number of important respects, including relative and absolute differences, the effect or total impact of inequalities in health and simple *versus* sophisticated techniques.

The measurement of relative or absolute differences. Most measures express the differences in the frequency of health problems between socioeconomic groups in relative terms: for example, the morbidity or mortality rate of the lowest groups as a percentage of that of the highest socioeconomic groups. One can also express these differences in absolute terms, such as the difference between the morbidity or mortality rates of the highest and the lowest socioeconomic groups.

Both perspectives are important: relative differences are usually more readily understood, but a 50% higher rate of a rare health problem may be much less important for the public's health than a 10% higher rate of a frequent health problem. Most of the relative measures can easily be transformed into absolute measures, and vice versa, and we argue that one should always look at both.

The measurement of an effect of lower socioeconomic status on morbidity or mortality or of the total impact of these inequalities in health on the health status of the general population. The main difference between these two perspectives is that the measures of total impact take into account not only the effect of decreasing socioeconomic status on health but also the extent of inequalities in socioeconomic status within the population: for example, by taking into account the size of the groups with lower socioeconomic status. The larger the extent of inequalities in socioeconomic status, the higher these measures of total impact will be. It is very much a matter of *a priori* choices whether one should take into account the extent of inequalities in socioeconomic status within the population. The size of the groups with lower socioeconomic status is largely outside the sphere of influence of public health policy, and this supports using a measure that focuses on the modifiable aspect: the effect of a lower socioeconomic status on health. Nevertheless, policy-makers address some features of the distribution of the population across socioeconomic groups (such as the income distribution), and this supports using the more comprehensive measures of total impact. We recommend using both types of measure and making a judgement based on a comparison of the results.

Simple versus sophisticated measurement techniques. The available measures differ substantially in degree of sophistication. Simple measures such as rate ratios or rate differences between a lower and a higher group have the advantages of easy calculation and straightforward interpretation and do not pose many restrictions on the data used in calculation. More specifically, the required measurement scale for the independent variable, socioeconomic status, is ordinal or even nominal only. The problem with these simple measures, however, is that they ignore parts of the available information. For example, the rate ratio measures do not take into account the morbidity or mortality rates of the socioeconomic status groups between the highest and lowest groups.

More sophisticated measures, particularly a number of regression-based measures, take more of the available information into account but only at the expense of greater complexity and more restrictions on the data used in calculation. Regression analysis *(111,112)* requires the socioeconomic status variable to be measured on an interval scale, which may be problematic, especially for occupation as an indicator of socioeconomic status. Perhaps a good recommendation is that policy-makers should always ask for simple measures but that researchers should always try to check the results obtained with simple measures against the results obtained with methodologically more refined measures.

Based on this analysis of the summary measures that have previously been applied, we have developed the threefold classification presented in Table 7.

Table 7. A systematic overview of possible summary measures for the magnitude of socioeconomic inequalities in health (the absolute versions of measures are in italics)

Degree of sophistication	Indices of effect	Indices of total impact	
		No inequalities = everyone has health of high socioeconomic status	No inequalities = everyone has health of average socioeconomic status
Simple	Rate ratio of lowest *versus* highest group	Population-attributable risk (%)	Index of dissimilarity (%)
	Rate difference of lowest versus highest group	*Population-attributable risk (absolute version)*	Index of dissimilarity *(absolute version)*
Sophisticated	Regression-based index of relative effect	Regression-based population-attributable risk (%)	Relative index of inequality
	Regression-based index of absolute effect	*Regression-based population-attributable risk (absolute version)*	*Slope index of inequality*

Twelve different summary measures

The three characteristics identified above can be combined into eight different classes of summary measures. Because, as will be shown below, the measures of total impact can be further subdivided in two groups based on their underlying assumptions, there are then 12 types of summary measure, each represented by a cell in Table 7. For each type, Table 7 mentions our recommended example. These recommendations are partly based on our evaluation of the adequacy of each measure and partly based on a desire to develop a set of indices that have clear interrelationships. For example, we ignore the pseudo-Gini coefficient *(113)* because this does not adequately take into account the hierarchical nature of the socioeconomic status-variable. We also ignore the concentration index *(38)*, partly because of a number of practical problems, such as the difficulties of calculating a confidence interval and of adjusting for confounding variables, and partly because it has been shown to be mathematically similar to the relative index of inequality, which is a regression-based measure and therefore relates clearly to the other regression-based measures in Table 7.

This systematic development of a set of possible measures identifies several measures that have not yet been applied (such as the absolute and sophisticated versions of the population-attributable risk). We briefly describe each of these possibilities, which are numerically illustrated in Boxes 2, 3 and 4. We recommend reading the main text together with these boxes, because they contain details of calculation that are essential for understanding the measures. We start with four measures of effect (Box 2).

1. Rate ratio of lowest versus highest socioeconomic status-group. There is substantial flexibility in the choice of the groups to be compared. For example, trends in inequalities in mortality in England and Wales have been studied by ratios comparing social class V to class I, by comparing classes IV+V to classes I+II or by comparing manual to nonmanual workers *(34)*. The choice of the groups to be compared should be a compromise between two conflicting requirements. The two groups should not be so extreme that the summary measure ignores most of the existing health inequalities and is sensitive to the idiosyncrasies of the two groups, but the two groups should not be so broad that the summary measure conceals the real extent of health inequalities in the population.

2. *Rate difference of lowest versus highest socioeconomic status-group.* The absolute equivalent to the ratio of the rates is the difference between the rates of two groups that are compared.

3. *Regression-based relative effect index.* A drawback of indices 1 and 2 is that they only take into account health inequalities between the two socioeconomic groups that are compared, and that they ignore all other health inequalities (either within the two groups or compared with groups that are excluded from the comparison). For that reason, regression-based indices have been developed that consider all socioeconomic groups separately and that assess how morbidity and mortality rates vary according to the socioeconomic status of these groups *(114,115)*. A convenient way of doing this is to apply a regression analysis in which morbidity or mortality rates (the dependent variable) are related to socioeconomic status (the independent variable). As mentioned before, this requires socioeconomic status to be measured on an interval scale, such as years of education or average income level in monetary units. For occupation this is problematic, although one might consider using one-dimensional scales. Depending on the specifications of the regression model, this approach produces a relative effect index, as in logarithmic or logistic transformations of the dependent variable, where regression coefficients translate into relative risks or odds ratios or a regression-based absolute effect index.

4. Regression-based absolute effect index. This type of index is used for regression analysis of untransformed morbidity or mortality rates. Usually, the absolute effect index can easily be transformed into a relative effect index (and vice versa) by relating it to the estimated rate of health problems in the reference category.

We continue with a number of measures of total impact, all related to a measure familiar to epidemiologists *(15–17)*, the population-attributable risk. The population-attributable risk and related indices are illustrated in Box 3.

5. *Population-attributable risk (%).* Although the population-attributable risk, sometimes also called the etiological fraction, is part of the standard repertoire of epidemiology, its application to the study of health inequalities is fairly recent *(103,116,117)*. This measure can be interpreted as the proportional reduction in overall morbidity and mortality rates that would occur in the hypothetical case that everyone experiences the rates of the highest socioeconomic group. It is calculated as the difference between the overall rate and the rate for the highest socioeconomic group, expressed as a percentage of the overall rate. The population-attributable risk not only reflects the morbidity and mortality rates of lower socioeconomic groups as compared with the highest socioeconomic group but also their population size: the larger the groups with high rates, the larger the potential reduction in overall rates is.

6. *Population-attributable risk (absolute version).* Multiplying the population-attributable risk by the overall rate yields a measure of the absolute reduction in the overall rate in the hypothetical case that everyone experiences the rates of the highest socioeconomic group.

7. *Regression-based population-attributable risk (%).* The simple calculation of the population-attributable risk is used at the price of ignoring the association between socioeconomic status and morbidity and mortality among the socioeconomic groups in worse health than the reference group. A more sophisticated index could be used – but has not yet been applied – that takes into account the association between socioeconomic status and morbidity and mortality across the entire social hierarchy. Just as with the effect indices, this sophisticated version of the population-attributable risk can be calculated by a regression analysis. The first step is to calculate the regression-based effect index (index 3 or 4), and the next step is to recalculate the population-attributable risk. This calculation is identical to that of the simple population-attributable risk, except that the reference rate is not the observed rate of the highest socioeconomic group but the predicted rate estimated for some high socioeconomic status value (such as 20 years of education). The corresponding interpretation is that this index measures the proportional reduction in overall morbidity and mortality rates that would occur in the hypothetical case that everyone experiences the rate that corresponds to 20 years of education or any other high socioeconomic status according to the regression model.

8. *Regression-based population-attributable risk (absolute version).* An absolute version of index 7 can be obtained by multiplying by the morbidity or mortality rate in the population as a whole.

Finally, we describe four alternative measures of total impact. These measures were first used in the study of trends in inequalities of mortality in England and Wales and were applied to take into account the changing distribution of the population over social classes and, more specifically, to take into account the decrease over time in the proportion of workers in unskilled manual occupations *(36,37,118,119)*. The calculation procedures that were developed account for the number of people in each social class. As a consequence, these measures, perhaps inadvertently, turned out to be measures of total impact. These measures are illustrated in Box 4.

9. *Index of dissimilarity*. The index of dissimilarity can be interpreted as follows: the percentage of all cases (for example, ill individuals or deaths) that has to be redistributed to obtain the same morbidity or mortality rate for all socioeconomic groups. The index of dissimilarity is larger if the groups with the highest and lowest rates are larger. This usually implies that the index of dissimilarity is larger if a relatively large part of the population is in the lowest and highest socioeconomic groups and relatively few people occupy intermediate positions. In other words, the index of dissimilarity is larger if inequalities in socioeconomic variables themselves are relatively large. The interpretation of the index of dissimilarity is subtly but significantly different from that of the population-attributable risk: whereas the index of dissimilarity reflects the extent to which the population distribution approaches the situation of an equal socioeconomic level for all, the population-attributable risk reflects the extent to which the population distribution approaches the situation of an equal high socioeconomic level for all: for example, postsecondary education or a specific high income level. The choice between the index of dissimilarity and the population-attributable risk largely depends on which perspective is considered most appropriate in a specific situation and may provide interesting material for discussion among policy-makers.

For example, if the average educational level of a population increases over time but the variation in education around the average remains equally large, the population-attributable risk would decrease, because more people had moved into the highest educational group, whereas the index of dissimilarity would remain unchanged (all other things being constant). Here one could argue that the population-attributable risk accurately reflects the fact that the upward shift in education allows more people to share in the favourable experience of highly educated people, which improves the health status of the population as a whole. Nevertheless, the index of dissimilarity accurately reflects the fact that the variation within the population has remained the same. Frequently, there is no clear preference for either way of looking at distributions of socioeconomic variables, and it is then highly informative to use the two indices complementarily.

10. *Index of dissimilarity (absolute version)*. Multiplying the index of dissimilarity by the overall rate yields a measure of the absolute number of cases that has to be redistributed to obtain equality in morbidity or mortality.

11. *Relative index of inequality*. A drawback of the index of dissimilarity is that it does not take into account whether high morbidity and mortality rates are found in the lower, the higher or intermediate socioeconomic groups. This was a reason to develop the relative index of inequality, a sophisticated measure that takes into account both the population size and the relative socioeconomic position of groups. It does so by regressing the morbidity or mortality rate of socioeconomic status groups on a very specific measure of their relative position: the proportion of the population that has a higher position in the social hierarchy. The resulting

figure can be interpreted as the ratio of the morbidity and mortality rates of those at the bottom of the social hierarchy compared with those at the top of the hierarchy, estimated based on the systematic association between morbidity and mortality and socioeconomic status for all groups. A large score on the relative index of inequality implies large differences in morbidity and mortality between high and low positions in the social hierarchy. This large morbidity and mortality difference can be attributed to:

- a large effect of socioeconomic status on morbidity and mortality, that is, a large score on the effect index; and
- large differences between high and low social positions in the socioeconomic indicator itself, that is, large variation in socioeconomic variables themselves.

With regard to the latter component, the interpretation of the relative index of inequality is analogous to that of the index of dissimilarity, and thus contrasts with measures similar to the population-attributable risk.

12. Slope index of inequality. The absolute equivalent to the relative index of inequality expresses the health inequality between the top and bottom of the social hierarchy in terms of rate differences instead of rate ratios. When the index of dissimilarity and relative index of inequality are presented to policy-makers, they must be properly interpreted. The problem with the relative index of inequality is that it is complex to interpret and easily misunderstood. The problem with the index of dissimilarity is that its interpretation (the percentage of cases to be redistributed to obtain equality) conveys the suggestion that health inequalities are a redistribution problem. This suggestion contradicts the message of health for all target 1, which advocates that inequalities should be reduced by improving the health of disadvantaged groups and not just by redistributing health problems from disadvantaged to advantaged groups.

Describing trends in health inequalities

Monitoring trends in health inequalities is essential for evaluating progress towards health for all target 1, which states that "by the year 2000, the differences in health status between countries and between groups within countries should be reduced by at least 25%, by improving the level of health of disadvantaged nations and groups". More generally, central to monitoring trends is the question of whether and how much health inequalities have been reduced.

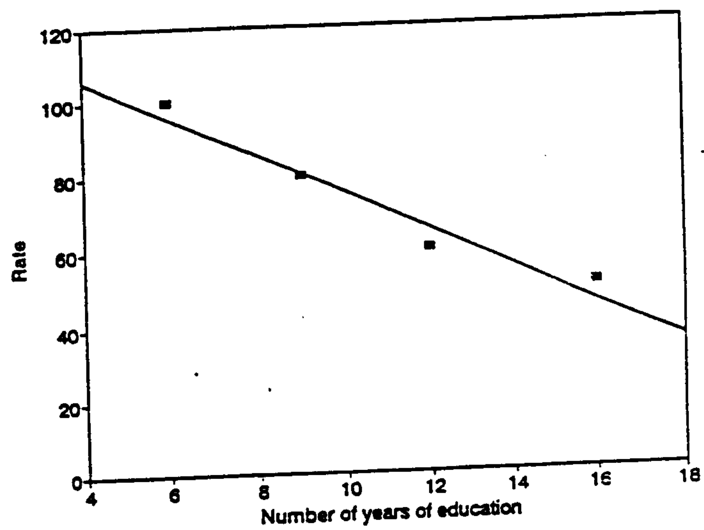
A detailed description of trends in health inequalities starts with the calculation of group-specific morbidity and mortality rates for each period separately (Box 5). One of the questions that should be examined is whether rates change systematically across the social hierarchy in each of the periods to be compared, so that the two periods can be compared by means of summary indices.

BOX 2. Summary indices of effect: simple and sophisticated

Using the data presented in Box 1, various measures can be calculated to summarize the size of the effect of educational level on the frequency of health problems. A simple measure is the rate ratio and rate difference of the lowest *versus* the highest educational group: 2.0 and 50 (cases per 10 000 population) in this example.

The calculation of a regression-based measure of absolute effect is illustrated in the figure below. Rates of health problems are related to the following measure of socioeconomic status: the number of years of education required to complete a specific educational level. In this example, the number of years corresponding to elementary education was 6 years. For higher levels, the numbers of years required were 9, 12 and 16. The frequency of health problems increases nearly linearly with decreasing number of years of education. The line in the figure is the regression line, which relates the rate of health problems (dependent variable) to the number of years of education (independent variable). Its slope has the value of –5.07, which implies that one additional year of education is estimated to be associated with a decrease in the rate of health problems by 5.07 cases per 10 000 population. In relative terms, this decrease associated with one additional year of education is 10% of the rate of health problems in the group with the highest level of education.

The regression model applied in this example is ordinary least-squares regression. It is often more appropriate to use a more sophisticated regression technique such as Poisson (log-linear) or logistic regression. If such techniques are used the effect estimates are given in relative instead of absolute terms, that is as the proportional increase in the morbidity or mortality rate for each less year of education or the proportional decrease in the rate of health problems for each additional year of education depending on the choice of the reference category. These relative effect measures can be transformed into absolute effect measures by relating them to the rate of health problems in the reference category.



BOX 3. Summary indices of total impact: the population-attributable risk

The simplest way to calculate the population-attributable risk is to subtract the rate in the reference category from the rate in the total population, and then to divide by the rate in the total population:

$$\text{Population-attributable risk} = \frac{81 - 50}{81} = 0.38 = 38\%$$

The population-attributable risk is estimated to be 38%, which implies that the total rate of health problems would be 38% lower if everyone experienced the rate of those with postsecondary education. Population-attributable risk is a relative measure, and its absolute equivalent can be obtained by multiplying by the rate in the total population (or by simply taking the numerator of the calculation given above):

$$\text{Population-attributable risk} \times 81 = 31 \text{ (cases per 10 000)} = 81 - 50$$

Although the simple calculation of the population-attributable risk does not immediately disclose this, the population-attributable risk reflects both the rate ratios of all the educational groups lower than postsecondary and their shares of the population. The larger the groups with the highest rate ratios are, the higher the rate in the total population will be as compared with the rate in the reference category, and thus the larger the population-attributable risk will be. This can easily be seen in the following alternative formula for the population-attributable risk:

$$\text{Population-attributable risk} = \frac{\Sigma pi(RRi - 1)}{\Sigma pi(RRi - 1) + 1}$$

with p = population share
RR = rate ratio
i = subscript denoting separate categories (such as educational groups).

In this example, this alternative formula would lead to the following estimate of population-attributable risk.

$$\text{Population-attributable risk} = \frac{(0.1 \times 0.0) + (0.2 \times 0.2) + (0.3 \times 0.6) + (0.4 \times 1.0)}{(0.1 \times 0.0) + (0.2 \times 0.2) + (0.3 \times 0.6) + (0.4 \times 1.0) + 1} = \frac{0.62}{1.62} = 0.38$$

Calculation of the regression-based version of the population-attributable risk starts by estimating the frequency of health problems that could be expected for the highest educational level, taking into account the linear association between level of education and morbidity rates over all educational groups. According to the regression model presented in Box 2, the rate of health problems that corresponds to 16 years of education (the educational level of the highest group) is 45 (cases per 10 000 population). The corresponding population-attributable risk is calculated as (81–45)/81, which gives a figure of 0.44 or 44%. Thus, approximately the same population-attributable risk is calculated when the association between level of education and rates over all educational groups is taken into account. The small difference between the simple population-attributable risk (0.38) and the regression-based population-attributable risk (0.44) is caused by the fact that the observed rate of health problems in the group with postsecondary education *(50)* is higher that the rate expected in this group on the basis of the regression model *(45)*. See also the figure in Box 2.

BOX 4. Summary indices of total impact: the index of dissimilarity and the relative index of inequality

The calculation of the index of dissimilarity is illustrated in the table below.

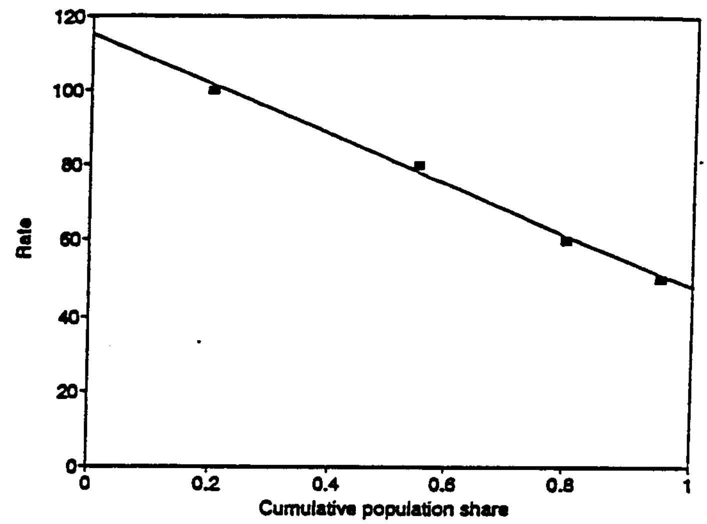
Educational level	Number of people	Rate of health problems	Number of people with health problems		Absolute difference |(3) – (4)|
			Observed	If there were equality	
	(1)	(2)	(3)	(4)	(5)
Postsecondary	10 000	50	50	81	31
Secondary, upper level	20 000	60	120	162	42
Secondary, lower level	30 000	80	240	243	3
Elementary	40 000	100	400	324	76
Total	100 000	81	810	810	152

Column 4 gives the number of cases that would occur in an educational group if that group had the same rate of health problems as observed in the total population (81 cases per 10 000 population). The difference between this number and the observed number of cases is given in column 5. It represents the number of cases that should be added or subtracted to get the same rate as in the total population. Summing that number over all educational groups gives a figure of 152. Since each case has been counted twice in that sum (one time for leaving a group and one time for entering another group), that total number of cases is divided by two. The resulting figure, 76, is the absolute version of the index of dissimilarity. It means that 76 cases have to be redistributed between educational groups to obtain the same rate (81 per 10 000) for each educational group.

The relative version, the index of dissimilarity proper, expresses the absolute number of cases to be distributed as a percentage of the total number of cases (810). This gives a figure of 9.4%. Thus, taking the index of dissimilarity calculation literally, complete equality in health could be attained by redistributing 9.4% of all cases.

The calculation of the absolute version of the relative index of inequality, the slope index of inequality, is illustrated in the figure below. By means of regression analysis, the rate of health problems is related to a measure of the relative position of educational groups. The slope index of inequality takes into account the population size of educational groups. More precisely, it is equated to the proportion of the population that has a lower position in the social hierarchy. For example, for the highest educational level, which comprises 10% of the population, that proportion ranges from 1.0 to 0.9, the average being 0.95. According to the same procedure, the score for lower educational levels is estimated to be 0.80, 0.55 and 0.20. The rates of health problems increase linearly with this measure. The fitted regression line has a slope of –67.4, which implies that, according to the regression equation, the health difference between those at the top of the social hierarchy (at 1) and those at the bottom (at 0) amounts to 67.4 (cases per 10 000 population).

The original version of the relative equivalent of the slope index of inequality, the relative index of inequality, is calculated as the quotient of the slope index of inequality and the overall mortality or morbidity rate *(37,119)*. To achieve consistency within our set of summary measures, we use a slightly modified calculation procedure. We first calculate the quotient of the slope index of inequality and the rate of health problems that is predicted for those at the top of the social hierarchy. This gives a figure of (67.4/47.2=) 1.43, which implies that, according to the regression equation, the rate among those at the bottom is 1.43 times higher than the rate among those at the top of the social hierarchy. We then express this as a rate ratio by adding 1, which gives a (modified) relative index of inequality of 2.43, indicating that the rate at the bottom is 2.43 times as high as the rate at the top.



This description should also provide a first impression of trends in health inequalities. The inspection of rate ratios for each period gives an impression of changes in health inequalities in relative terms, whereas the inspection of rate differences is required to examine changes in absolute terms. These two perspectives may produce different results: health inequalities may increase in relative terms but decrease in absolute terms, whereas in some situations just the opposite may occur.

A new element is the description of trends in morbidity and mortality rates in individual socioeconomic groups. This assessment is greatly facilitated by graphical analysis, which shows how a decrease or increase in inequalities in health has resulted from trends in individual socioeconomic groups. Several situations may occur. For example, a decrease in

BOX 5. Basic description of trends in health inequalities

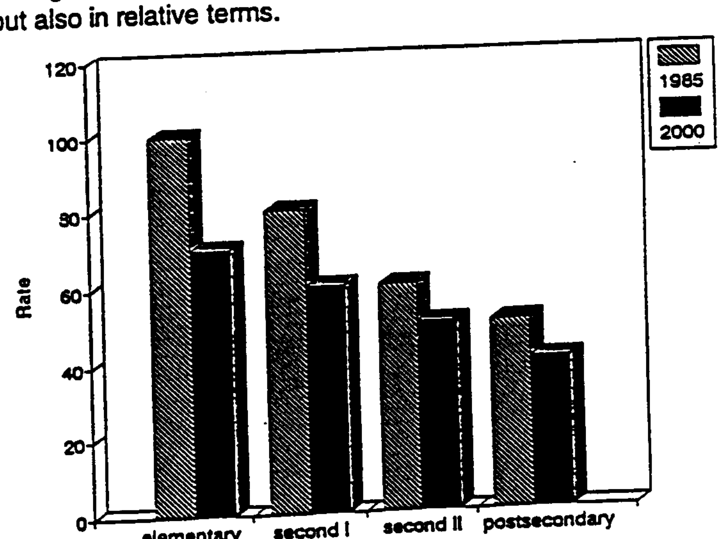
The table below describes a fictitious example of two observations on socioeconomic inequalities in health, one in 1985 and one in the year 2000. The situation described for 1985 is identical to the one used in Boxes 1–4. The situation described for 2000 is characterized by a lower overall rate of health problems and by a lower rate in all educational groups.

Educational level	Population share		Rate of health problems		Change in rates	
	1985	2000	1985	2000	Relative (2000 / 1985)	Absolute (2000 – 1985)
	(1)	(2)	(3)	(4)	(5)	(6)
Postsecondary	0.10	0.20	50	40	0.80	10
Secondary, upper level	0.20	0.30	60	50	0.83	10
Secondary, lower level	0.30	0.30	80	60	0.75	20
Elementary	0.40	0.20	100	70	0.70	30
Total	1.00	1.00	81	55	0.68	26

The figure below, based on columns 3 and 4 of the table, shows that in 2000 and in 1985 the frequency of health problems decreases consistently with increasing educational level. The size of inequalities has clearly decreased in absolute terms, but it cannot be decided from this visual inspection whether inequalities also decreased in relative terms.

The last two columns of the table describe the trend in the rate of health problems in each socioeconomic group. Between 1985 and 2000, the overall rate decreased from 81 to 55 cases per 10 000 population. In relative terms, this decrease is 32% (1–0.68) and in absolute terms it is 26 cases per 10 000 population. The largest decrease occurred for the group with elementary education, not only in absolute terms but also in relative terms.

This basic description cannot demonstrate whether the decrease in inequalities in health has been sufficiently large to meet a specific target such as the 25% reduction target for health for all. Nor is it clear what the effect would be of taking into account the changing population distribution. This evaluation requires the use of summary indices, which are illustrated in Box 6.



health inequalities in both absolute and relative terms may result from three different situations:

- increasing morbidity and mortality rates in all socioeconomic groups but smaller increases in lower groups;
- decreasing morbidity and mortality rates in all socioeconomic groups but larger decreases in lower groups; and
- increasing morbidity and mortality rates in higher groups and decreasing rates in lower groups.

The first of these situations is clearly not in agreement with the concern expressed in the first health for all target, that a decrease in health inequalities should be achieved by reducing the morbidity and mortality rates in the lower socioeconomic groups.

Summarizing trends in the size of health inequalities

A basic description of trends in health inequalities is often not sufficient to decide whether a 25% reduction of health inequalities (or any other target) has been achieved, so that summary indices are needed to quantify the change in the size of health inequalities. Box 7 illustrates the use of the simple summary indices that were recommended in the second section of this chapter. This section discusses the relative weight that should be given to the results from these various indices.

Trends in health inequalities can be evaluated in both relative and absolute terms. Trends in health inequalities that are expressed in absolute terms are sensitive to the trend in overall morbidity and mortality, whereas trends expressed in relative terms are not. Whether relative or absolute measures are more important depends, among other factors, on the health indicator under consideration. As a general rule, a public health policy perspective advocates looking at absolute differences, because these are what really matters for the public's health, whereas relative differences may help to explain the causes of (changes in) health inequalities. A strong upward trend in the overall average morbidity or mortality rate may make it virtually impossible and a strong downward trend may make it very easy to reach a 25% (or other) absolute reduction, and to achieve this by improving the level of health of disadvantaged groups.

When measures of total impact are used, the monitoring takes into account changes in the distribution of the population over the socioeconomic categories. Whether or not these changes should be taken into account again very much depends on the perspective chosen. If the population distribution across socioeconomic groups is considered to be completely outside the influence of public health policy, it may be preferable to ignore any changes that have occurred in this distribution. One may then be more interested in narrowing or widening the health gap between specific socioeconomic groups, irrespective of their size. On the other hand, it can be argued that public health policy can or should address some aspects of the population distribution across socioeconomic groups. The distribution of the population across income groups (the income distribution) could be a good example. If this stance is taken, a measure of total impact would be most suitable.

Among measures on the total impact of socioeconomic inequalities on the morbidity or mortality rate of the population, the choices are between the population-attributable risk (and its sophisticated equivalents) or the index of dissimilarity (and its sophisticated equivalents, such as the relative index of inequality). The choice of either type of index largely depends on the perspective chosen on the changes in the distribution of the total population over socioeconomic groups. If one wants to take into account all changes in the extent of variation of socioeconomic levels around the population average, the index of dissimilarity is most appropriate. The population-attributable risk is more appropriate if the emphasis is on the increase or decrease of differences in socioeconomic levels as compared with some fixed higher level. The conceptual difference between the two measures may seem subtle, but the examples presented in Box 6 and in part 3 show that estimates of trends in health inequalities can strongly depend on the measure chosen.

It is advisable to use both simple and sophisticated indices. The virtue of simple indices is that policy-makers can readily understand their calculation and interpretation. The price of their simple interpretation is that they ignore part of the available information on inequalities in health (such as the morbidity and mortality rates of intermediate groups). Sophisticated indices are therefore needed to check whether they show the same trends in health inequalities. If not, an inspection of the detailed patterns is needed to determine whether the reliance on simple indices has biased the estimate of the extent to which health inequalities changed.

BOX 6. Summary measures for trends in the magnitude of health inequalities

The trend in health inequalities in the fictitious example of Box 5 is summarized in the table below by means of simple indices.

Inequality measure	Year		Relative change
	1985	2000	(2000/1985)
Rate ratio[a]	2.00	1.75	0.75
Rate difference[a]	50	30	0.60
Population-attributable risk[b]	0.38	0.27	0.71
Population-attributable risk x overall rate[b]	31	15	0.48
Index of dissimilarity	0.094	0.164	1.74
Index of dissimilarity x overall rate	7.6	9.0	1.18

a Elementary *versus* postsecondary education.

b Postsecondary education as reference category.

Six summary indices are presented. The rate ratio and the rate difference are based on a comparison between the highest and lowest educational level. The rate ratio decreased to 75% of its initial value (0.75=(1.75–1)/(2.00–1)). The rate difference decreased even more because the overall rate of health problems declined. The rate difference in the year 2000 is only at 60% of its value for 1985. This 60% estimate can be differentiated into two components: the decrease in the health disadvantage of the elementary level over the postsecondary level (the rate ratio) and the decrease in the rate of the postsecondary level itself.

The population attributable risk showed a decrease in the total impact of the observed health inequalities. Calculating the population-attributable risk according to the simple method illustrated in Box 3 yields a value of 0.27 for the year 2000. This population-attributable risk is only 71% of the 1985 value. Even in absolute terms, the population-attributable risk decreased to less than one half of its initial value. The fact that the reduction is larger in terms of population-attributable risk than in terms of rate ratios or rate differences can be explained as follows: between 1985 and 2000 not only did health inequalities between high and elementary educational levels decrease (as reflected in the rate ratio and difference) but the share of the population with only elementary education decreased as well. As shown in Box 5, 20% of the population has elementary-level education in 2000 as compared to 40% in 1985.

According to the index of dissimilarity, the total impact of health inequalities increased during the monitoring period. Calculating the index of dissimilarity according to the method illustrated in Box 4 yields a value of 0.164 (16.4%) for the year 2000, which is larger than the 1985 value. Even in absolute terms, the index of dissimilarity for the year 2000 is larger than for 1985. The fact that the index of dissimilarity increased can be explained as follows: although health differences between educational levels decreased during the monitoring period, this was compensated by an increase in the variation in educational levels within the total population. As shown in Box 5, in 1985 a large part of the population was concentrated in the lowest educational levels whereas, in 2000, the population spread from high to low levels was much more even. This implied an increase in the average educational level (contributing to the decrease in population-attributable risk) but also a larger variation in educational levels (contributing to the increase in index of dissimilarity).

EVALUATING DATA PROBLEMS

A wide variety of data problems can potentially introduce serious bias in the measurement and monitoring of health inequalities *(120)*. It is therefore essential to check for data problems when health inequalities are estimated. This chapter provides a checklist of common data problems that can strongly bias estimates of health inequality. We first discuss problems that could affect the measurement of health inequalities at one moment in time, and then we discuss additional data problems that should be considered when trends in health inequalities are studied over time.

Problems with measuring health inequalities at one point in time

Data problems can be classified into three groups *(15)*: problems of external validity (the study population is not representative of the total population of interest), problems of internal validity (leading to biased estimates of the association between socioeconomic status and health in the population covered by the study), and problems of precision (too much random error).

Problems of external validity

Several problems can affect the representativeness of the population covered by the study for the total population of interest.

1. Restriction to specific age and sex groups

The problem. Some studies are confined to men or, less often, to women. Most studies on socioeconomic inequalities in health cover only part of the entire age range (for example, 15–70 years). Since the magnitude of these inequalities can vary strongly by age and sex *(18–23,70,115)*, the results from one age and sex group cannot without reserve be generalized to the total population.

What can be done. Little. Extrapolation to other age and sex groups is highly speculative. As a general rule, socioeconomic inequalities in mortality are smaller among women than among men, and after the age of about 35 years they diminish with increasing age. This general rule, may be used to speculate on the possible size of socioeconomic inequalities in mortality (but not in morbidity) among age and sex groups that are not included in the data.

2. Restriction to specific subpopulations

The problem. Some data sources are confined to specific subpopulations such as civil servants or workers in specific companies, such as the Whitehall study *(26)* and one of the few mortality studies from Belgium *(121)*. When no data are available at the national level, there are often relevant studies in specific regions or cities. Unfortunately, however, the magnitude of the socioeconomic inequalities in health can vary between regions *(23,31,44)*, degree of urbanization *(23,122)* or, more generally, from one subpopulation to other. As a consequence, the results from one subpopulation cannot without reserve be generalized to the total population of interest.

What can be done. Little. If there are various studies on different subpopulations, estimates of socioeconomic inequalities in health from all these studies should be compared to see whether they consistently point to the same magnitude of health inequalities. Whenever possible, any available data should be used to assess whether it is likely that the health inequalities within the studied subpopulation are larger or smaller than the health inequalities in the total population of interest. When the study is restricted to economically active people (for example, employed men enrolled in a health and social security fund), a comparison of their morbidity or mortality rates to the rates in the general population can show to what extent the healthy worker effect occurred and, as a consequence, how large the bias would be if this healthy worker effect is larger among some occupational groups than among others *(91,123)*.

3. Exclusion of specific subpopulations

The problem. Many data sources exclude specific and relatively small groups such as foreigners, the institutionalized population or people not entitled to social or health insurance. The effect of their exclusion on the estimates of health inequality is highly unpredictable but could be substantial if the excluded group has both distinctive socioeconomic characteristics and a characteristic health profile, such as immigrant groups *(107,124,125)*.

What can be done. Indications of the potential magnitude of bias may be obtained by assessing the extent to which the excluded group is similar to the total population of interest in terms of health and socioeconomic indicators. More confidence should be given to estimates of socioeconomic inequality in health for age and sex groups when the excluded subpopulations make up only a small part of the total population.

4. Nonresponse

The problem. Health interview and similar surveys face substantial (more than 20%) nonresponse in such countries as Germany and the Netherlands *(126)*. The exclusion of nonresponders leads to biased estimates of the size of health inequalities if nonresponse is associated with socioeconomic status and, in addition, if nonresponse is associated with health status (given a certain socioeconomic status).

What can be done. Indications of the potential magnitude of bias may be obtained from any data on characteristics of nonrespondents. For example, data on their place of residence may indicate whether nonrespondents are a highly selective group in terms of socioeconomic status, and data on the reasons for nonresponse may provide indications of their morbidity levels.

Problems of internal validity

Several problems can bias the estimates of the magnitude of socioeconomic inequalities in health within the population covered by the study.

1. Health interview surveys: reporting behaviour

The problem. A fundamental limitation of the use of health interview survey data is that they depend on reporting by subjects. Self-reports of health problems not only depend on the objective presence of health problems and the subjective experience of these problems but also on factors related to their reporting, such as diagnoses previously made by physicians, the subject's interpretation of the diagnostic terms used in the survey questionnaire, the failure to recall health problems and the propensity to complain *(60,127)*. If, as some studies observed *(128,129)*, respondents with a low educational level are less likely to report a chronic condition, this would underestimate the magnitude of inequalities in the prevalence of chronic conditions.

What can be done. Interpretation of results should take into account the possibility that part of the observed inequality in self-reported morbidity is related to differential reporting tendencies. The direction and magnitude of bias is usually only a matter of speculation. The measurement of health inequalities should rely on various health indicators and attach more importance to indicators based on questions that leave less room for reporting bias.

Special caution is required for indicators of illness behaviour, such as restriction of daily activities and confinement to bed, because inequalities in illness behaviour may reflect inequalities not only in the presence of health problems but also in such factors as the social valuation of illness behaviour and financial pressures to continue one's daily activities. Socioeconomic inequalities in the self-reported prevalence of chronic conditions should be checked, if feasible, against prevalence data on the same people from disease registries or registers or against data from the survey that are based on a battery of diagnostic questions on symptoms of a specific disease, such as chronic obstructive lung disease or ischaemic heart disease.

2. Unlinked mortality studies: numerator and denominator bias

The problem. Unlinked cross-sectional mortality studies combine data from two different sources of information that cannot be linked at the individual level: the vital registry gives the numbers of deaths by occupation, and the population census gives the corresponding numbers of the population by occupation. In many countries, the validity of this type of information is likely to be compromised to a varying extent by the poor comparability between the measurement of socioeconomic status on the death certificate and that in the population census.

The consequences for France are illustrated in Fig. 8 *(23)*. The left part of the figure, which is based on a national longitudinal study, shows a consistent association between mortality rates and occupational status: the lowest rates are observed for upper- and middle-level employees and the highest rates are observed for semiskilled and unskilled labourers. When unlinked cross-sectional data are used, however, the resulting pattern shows two major inconsistencies that are probably caused by an invalid registration of occupations in death certificates.

- First, unskilled labourers have a lower death rate than semiskilled labourers. Their low death rate is probably caused by "promoting the dead", that is, the relatives' tendency to avoid mentioning a low-status job for the deceased (such as, in this case, the French term *manoeuvre*). A consequence is that the number of deaths among people with low-status jobs is underestimated.

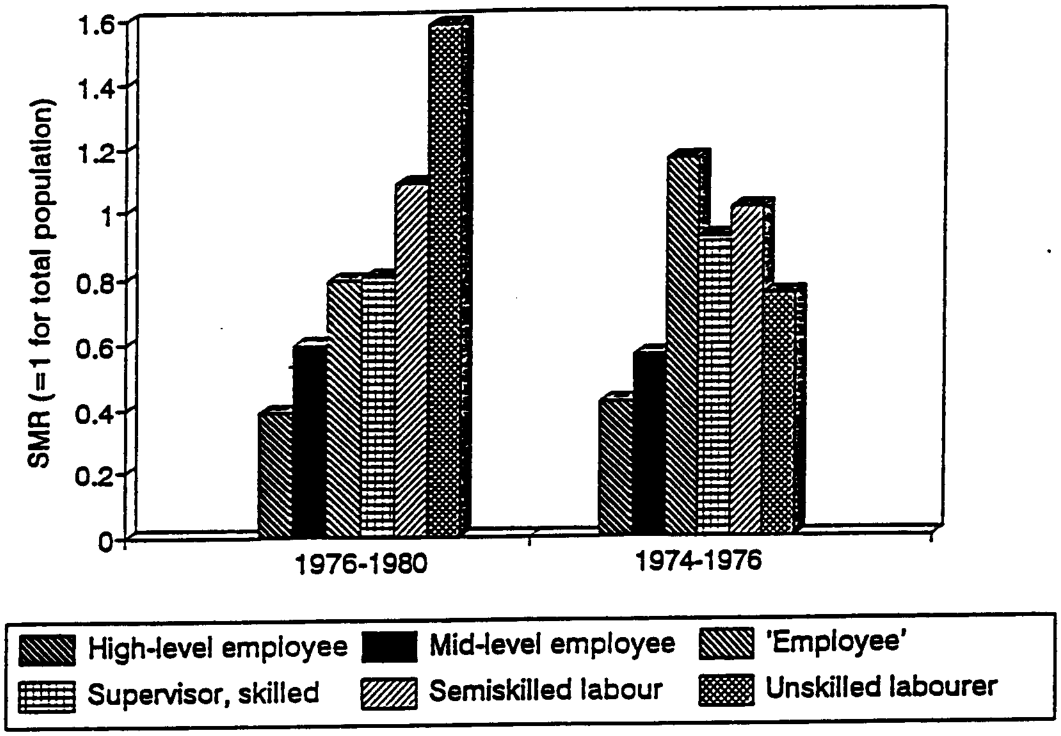
- Second, the mortality rate for lower-level employees is higher than for any other social class. This overestimation is probably caused by the frequent use of vague occupational descriptions in death certificates. Employees is a very popular term in France, and its frequent use overestimates the number of deaths among low-level employees.

What can be done. A global impression of the magnitude of numerator and denominator bias can be obtained by investigating the frequency of such vague occupational titles as labourers and employees in, respectively, death certificates and population census records. A more precise estimate of the magnitude of numerator and denominator bias would be possible when the occupation titles as registered in death certificates for a sample of deceased people are compared with the occupations of the same people as registered in the population census. For example, a study in Switzerland showed in this way for which occupational classes the mortality rates were likely to be grossly over- or underestimated *(130)*.

3. Ecological studies: the ecological fallacy

The problem. Studies that compare the morbidity and mortality rates of areas with low socioeconomic status with those of areas with high status are called aggregate level or ecological studies. These studies are often carried out to obtain some indications of the association between socioeconomic status and health when such information is not available at the individual level. In that case, one has to assume that the individual-level association can be inferred from the observed aggregate-level (area-level) association. Nevertheless, this cross-level inference is problematic *(131,132)*. One problem is that the socioeconomic status of an area is usually so strongly correlated with other factors (such as housing or environmental conditions) that it may be nearly impossible to separate the effect of each variable. A more fundamental problem is that the socioeconomic status of an area represents a different phenomenon than the socioeconomic status of individual people. For example, a high unemployment rate not only means that a relatively large proportion of the inhabitants of an area face unemployment but also implies that the community at large is economically disadvantaged and that the total working population faces on average a larger risk of becoming unemployed.

Fig. 8 Social class differences in mortality in France among men 45–49 years old. Comparison of estimates from longitudinal data for 1976–1980 and unlinked cross-sectional data for 1974–1976



Source: calculated from data presented in Desplanques (23).

What can be done. The observed health inequalities between areas cannot be used to estimate the precise magnitude of health differences at the individual level. They may be used, however, to obtain evidence of the effects of socioeconomic disadvantage on health. The quality of this evidence improves to the extent that the distinguished areas are smaller and socioeconomically more homogeneous. For that reason, studies from various countries have attempted to distinguish areas with a mean population size of less than 10 000 inhabitants *(44,45,47,133–135)*. Also, ecological studies should stick to the type of inequality indices proposed in the previous chapter. An association measure often used in ecological studies, the product-moment correlation, is less adequate because it does not express how much larger mortality rates are in nonaffluent areas than in affluent areas *(131,132)*.

4. Occupation: nonhierarchical classification

The problem. Although occupational classifications are seldom strictly hierarchical, most classifications permit an approximate ordering of people from high to low positions on the occupational hierarchy. Nevertheless, even this is hardly possible in some classifications, for example, with the distinction between the eight major groups of the International Standard Classification of Occupations *(136)*:

00–01	Professional, technical and related workers
02	Administrative and managerial workers
03	Clerical and related workers
04	Sales workers
05	Service workers
06	Agricultural etc. workers
07–09	Production and related workers, transport, labourers
10	Members of the armed forces

Mortality rates in Spain and the former Yugoslavia are only available for (a national version of) the classification comprising these broad groups *(40,43)*. The nonhierarchical nature of this classification implies that its use does not enable the researcher to demonstrate the precise extent of health differences across the social hierarchy.

What can be done. If a large part of the population is not divided into clearly hierarchical groups, the measurement of health inequalities should focus on the few groups that do have a clearly distinct socioeconomic status. When the eight major groups of the International Standard Classification of Occupations are distinguished, a comparison may be made between groups 00–01–02 and 07–09, which are at the high and low ends of the occupational hierarchy. Simple indices such as the rate ratio or the population-attributable risk are most suited for this because they can be focused on specific groups. If education and income are also available for the same data set, occupation does not have to be used as an indicator of socioeconomic status.

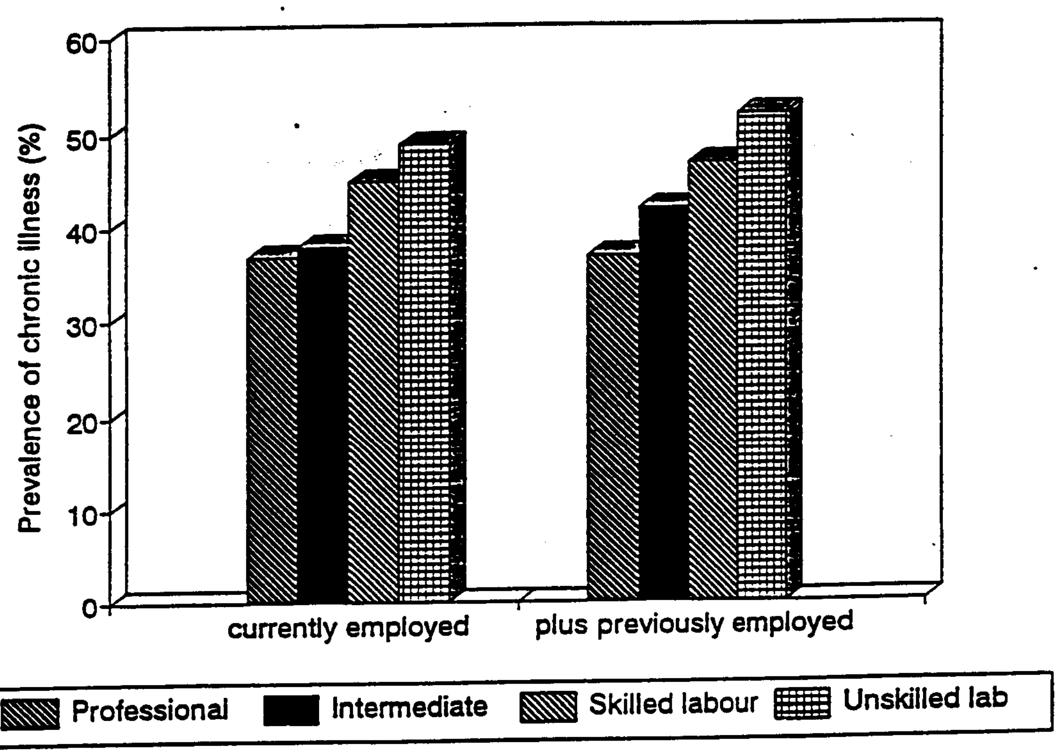
5. Occupation: the last occupation of economically inactive people is not known

The problem. Even though economically inactive people by definition do not have an occupation, in principle most of them could be assigned a place in the occupational hierarchy based on their last or longest held occupation. Information on previous occupations is often lacking, however, and economically inactive people then have to be excluded from the measurement of the association between health and social class. Economically inactive people are often less healthy on average, so their exclusion decreases the observed morbidity and mortality rates. As the economically inactive people often originate predominantly from the lower social classes, the morbidity and mortality rates of these classes are more underestimated than those of the upper classes. As a result, the exclusion of economically inactive people could substantially underestimate the magnitude of health differences between low and high occupational classes.

This point is illustrated in Fig. 9 for Norway *(91,123)*. A consistent social class gradient is observed among currently employed men (left part), but these differences are smaller than those observed when previously employed men are included and classified according to their last occupation (right part). For example, the ratio of the prevalence rates of the lowest to the highest social class increases from 1.32 in the first case to 1.41 in the second case.

What can be done. If the occupation of nearly all economically inactive people is unknown, measurement of the magnitude of health inequalities must be restricted to the economically active population, and the resulting estimates are likely to be smaller than those for the entire population of interest. Indications of the likely extent of underestimation can be obtained from a detailed description of the inactive population, including its size, its morbidity or mortality level and, if possible, its average socioeconomic status according to education and income data. In some situations, the analysis may need to be restricted to men and to age groups in which the share of inactive people in the total population and their share in the total number of diseases and deaths is relatively small. In longitudinal mortality studies, the bias related to exclusion of those who are economically inactive at the start of the follow-up can be reduced substantially by excluding the first five years of follow-up *(137)*.

Fig. 9 Social class differences in the prevalence of chronic illness in Norway among men 25–66 years in 1983. Comparison of estimates for currently employed men only and for currently and previously employed men



Source: Dahl *(123)*.

6. Education: large categories

The problem. Some data sources fail to distinguish between lower and intermediate levels of education, with the result that more than half the subjects are combined into a category such as "lower-secondary education or less". In the United Kingdom, for example, everyone without "higher qualifications" is often combined into one group that comprises a large part of the population; among elderly people it even comprises more than two thirds of the population *(22,82)*. In such cases, estimates of health inequalities using education as an indicator of socioeconomic status inevitably have to be restricted to a comparison between those with high education and all others with lower levels of education.

What can be done. Alternative measures of education with a more even distribution, such as the age at leaving school, may be considered. Otherwise, the analysis could be restricted to younger age groups, at least if the share of those with low education is relatively small in these groups. The measurement of health inequalities may be restricted to the use of simple indices that clearly express to which educational levels the comparisons are restricted: rate ratios (high educational level *versus* all other) and the population-attributable risk (the high educational level as the reference group).

7. Occupation, education or income not known

The problem. The occupation, education or income of a large number of people is often unknown. The most frequent problem with occupation is that the previous occupations of economically inactive people are not known. The other frequent problem refers to income: in many health interview surveys a large proportion of the respondents refuse to report their income level, even if the question on income levels only asks about broad income categories. The extent to which excluding these subjects leads to biased estimates of health inequality varies but may be substantial.

What can be done. Indications of the precise magnitude of bias can be obtained from a detailed description of the unknown group, including its population size, its morbidity or mortality level and, if possible, its socioeconomic status according to other socioeconomic indicators. If this group is not very large and it is close to the population average in terms of both socioeconomic status and morbidity and mortality, the unknown group probably does not present a major data problem. If there is strong evidence that the unknown group comes from a specific socioeconomic group, for example, if most subjects with education unknown come from the group of no education completed, the groups could be combined.

Problems of precision

The problem. The morbidity and mortality rates by socioeconomic group are subject to large random fluctuation in many studies because few cases (deaths or respondents who report health problems) are observed in each socioeconomic group. Because of this random fluctuation, estimates of health inequality may be very uncertain.

Table 8 illustrates this with an example from the Kaunas-Rotterdam Intervention Study, a project initiated by WHO consisting of two parallel epidemiological studies in the cities of Kaunas, Lithuania and Rotterdam, the Netherlands *(138)*. Both studies include a longitudinal mortality follow-up of 2452 and 3365 men, respectively. Death rates increase with lower educational level. Nevertheless, because the total numbers of deaths observed are small (303 and 350 respectively), 95% confidence intervals around the rate ratios are large. The rate ratios for Kaunas are not statistically significant: it cannot be demonstrated with 95% confidence that they differ from 1. The rate ratios for Rotterdam are statistically significant, but the confidence interval is so large that the precise magnitude of inequalities in mortality remains highly uncertain. For example, these wide confidence intervals show that there is a 5% probability that the death rate for people with elementary education is only 13% higher or less than the rate for people with postsecondary education instead of the observed 57%.

Table 8. Association between mortality and education level in Kaunas and Rotterdam among men 45–59 years in 1972–1982

City	Education level	Rate ratio	95% confidence interval
Kaunas	Postsecondary	1.00	
	Secondary	1.27	0.93 – 1.73
	Elementary	1.29	0.92 – 1.80
Rotterdam	Postsecondary	1.00	
	Secondary	1.49	1.06 – 2.10
	Elementary	1.57	1.13 – 2.18

Source: Bosma *(138)*.

What can be done. The precision of estimates of health inequality can be increased in several ways. A general strategy for treating health interview survey data could include:

Selection of surveys. Preference is given to fairly large surveys: at least 5000 respondents and, if possible, more than 10 000 respondents.

Selection of health indicators. Preference is given to indicators of frequent health problems. Examples are measures on the prevalence of all chronic conditions (instead of a few conditions only) and on the prevalence of less-than-good perceived health (instead of the rarely mentioned very poor health only).

Classification of socioeconomic variables. Small socioeconomic groups with similar socioeconomic status are combined into broad groups. When the magnitude of inequalities is measured by the population-attributable risk, one should avoid using a small reference group comprising less than about 20% of the population, such as professionals, or those with university education.

Confidence interval estimates. Confidence intervals are estimated to quantify the potential role of random fluctuation *(139)*. The example given above shows that, to assess the magnitude of health inequalities, it is not sufficient to test the statistical significance of health inequalities. Such a test informs about the likelihood that the observed inequalities are attributable to random fluctuation, but in case of a statistically significant association it does not inform about the remaining uncertainty on the precise magnitude of health inequalities.

Problems with measuring trends in health inequalities

When the magnitude of health inequalities is compared at two moments of time, inequality estimates for each period should be evaluated for the data problems discussed above. In addition, estimates of the changes in the magnitude of health inequalities should be evaluated against three new data issues: the comparability of data through time, the beginning and length of the monitoring period and the precision of estimates of change in health inequalities.

Comparability of data

As estimates of health inequality are very sensitive to data characteristics, the data for the different points in time should be as comparable as possible. This comparability is required for the study design, the delimitation of the study population, the measurement of mortality and morbidity, and the measurement and classification of socioeconomic variables. Three points merit special attention.

- Because the magnitude of socioeconomic inequality in health can vary strongly by age and sex, estimates of health inequality should refer to exactly the same age and sex groups. For monitoring purposes, it is generally not useful to follow one single birth cohort through time. Because a birth cohort grows older and because most inequalities in mortality and morbidity diminish with increasing age, the follow-up of a single birth cohort may show diminishing inequalities even when inequalities in the general population of all ages are increasing.

- The study of changes in the magnitude of health inequalities requires high comparability of the socioeconomic classifications for the points in time compared. This applies especially when the comparison is made by using simple summary indices such as the rate ratio and the population-attributable risk. For example, when the population-attributable risk is used, the same educational level in both points in time or, in the case of income, the reference group should be the same in terms of real income. The comparability of the socioeconomic indicators can be determined globally by comparing the observed population distribution among the two points in time. Are the changes in the distribution of the population over the socioeconomic categories in the expected direction? Are changes not larger than could be regarded as plausible? In most countries in Europe, the proportion of, among others, unskilled labourers and those with low education may be expected to decline.

- Changes in the occupational composition of social classes in the Registrar General's social class scheme are one of the main reasons why some authors doubt the validity of the data that suggest an increase in social class mortality differences in England and Wales since 1945 *(33)*. Nevertheless, the comparability of socioeconomic classifications does not imply that they should be identical for both points in time. Classifications often need to be adapted to changing circumstances. For example, because of technological change, some previously skilled occupations may become jobs that consist of simple machine operation, whereas new skilled occupations are emerging, such as those related to automation.

- Because the magnitude of inequalities observed for a morbidity indicator depends on the precise survey question, that survey question should be as comparable as possible for the two points in time. The comparability of the health indicator measurement can be determined globally by comparing the observed overall prevalence rates among the two points in time: are changes in these rates, if any, in the direction expected? Are changes not larger than could be regarded as plausible?

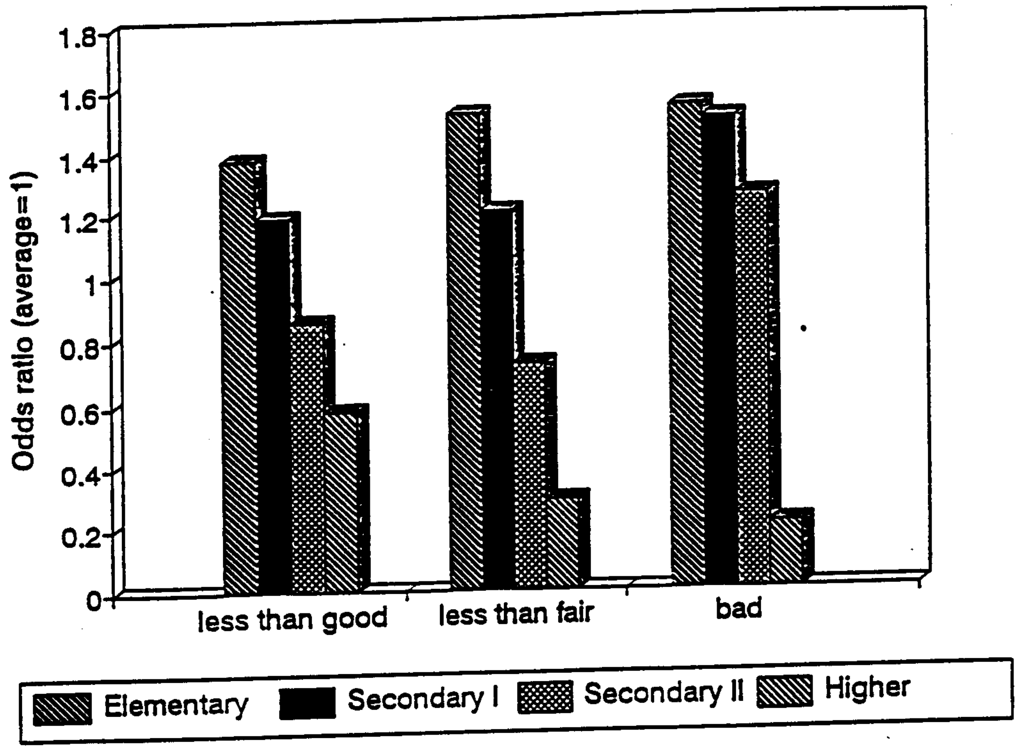
This sensitivity of the magnitude of health inequalities to the precise measurement of morbidity is illustrated in Fig. 10 using data from the Netherlands *(140)*. Large differences between educational groups are observed in the prevalence of less-than-good perceived general health (left graph). Even larger differences are observed for a more restrictive cut-off point (less than fair, middle graph). Thus, larger inequalities are observed when the survey question restricts the analysis to people with more serious health problems. When an even more restrictive cut-off point is chosen and the analysis is confined to the 2.4% of those surveyed reporting poor health, the observed inequalities take the form of a contrast between a low prevalence rate for the highest educational group and an equally high prevalence rate for all other groups.

Beginning and length of the monitoring period

The two points in time to be compared can often be chosen flexibly: for example, when data are available from a continuous health interview survey covering a long period. More commonly, however, the periods to be compared are determined by the availability of data for only a few periods, such as the years in which population censuses were held. In such cases, the adequacy of the chosen time period should be evaluated using two criteria.

First, the data for the beginning of the monitoring period should not be too old. Since the magnitude of health inequalities may sometimes change substantially in a period of about 10 years *(18,23,34–38,52,114,134,137,141–143)*, the difference between the intended and the real start of the monitoring period is preferably less than five years.

Fig. 10 Differences in perceived general health by educational level in the Netherlands among men 15 years and older 1989–1991. Comparison of estimates for different cut-off points.



The overall prevalence rates are 17.8% for "less than good", 69.% for "less than fair" and 2.4% for "bad".

Source: Leclerc et al *(113)*.

Second, the monitoring period should be sufficiently long to detect changes in the magnitude of health inequalities. About ten years is required to find substantial changes in the magnitude of health inequalities, although radical changes (such as in the countries of central and eastern Europe and the newly independent states in the early 1990s) and quickly responsive health indicators (such as mortality due to infectious diseases, traffic accidents or suicide) may justify a shorter monitoring period. A shorter monitoring period would also be justified when an intervention programme is launched and when the effects of this programme on (the determinants of) morbidity or mortality are expected to become visible within less than a decade.

Precision of the estimates of change

The role of random fluctuation should receive extensive attention when two moments in time are being compared. Demonstrating with sufficient precision a decrease in health inequalities of a magnitude of the 25% reduction that is targeted requires that the magnitude of health inequalities at each moment in time be measured with as small confidence intervals as possible. This condition is usually met in most national mortality studies but is more difficult to meet in, for example, health interview surveys. Therefore, much effort should be made to obtain estimates of health inequality that are very precise. The question of whether random fluctuation can explain the observed trend in the estimates of health inequality can be assessed by comparing the two health inequality estimates together with their respective confidence intervals *(139)*.