Part Six

DATA COLLECTION FOR INTENSIVE MORTALITY STUDIES
In recent years, interest in mortality differentials in developed countries has revived. There are two main reasons for this concern. The first reason is the observation in the post-war period in some developed countries of a stagnation or even an increase in male mortality. This unfavourable change has taken place simultaneously with an expansion in health care expenditures and has led to such concepts as "the diminishing return from health investments" and "the medical nemesis." The second reason is the general threat of ecological catastrophes caused by side-effects of technological innovations, especially from new chemicals and atomic energy.

The study of mortality differentials in developed countries today therefore serves to identify high-risk groups or unfavourable trends, which in turn indicate areas in which preventive interventions are likely to be fruitful. Valuable supplements to such mortality studies are studies of health factors other than mortality, such as cancer incidence, sterility, spontaneous abortions and congenital malformations, and studies of the effectiveness of treatment and prevention strategies.

All mortality studies share certain traditional elements, and well-tried techniques still have a role to play; but during the 1970s important methodological innovations were introduced into epidemiological studies of occupational risks, smoking, side-effects of drugs etc. This chapter gives a short description, illustrated with examples, of each of the methods currently available for the study of mortality differentials.

A. DETECTION OF MORTALITY DIFFERENTIALS BY PLACE AND TIME FOR POPULATION GROUPS

First indications of areas where health intervention may be fruitful can often be obtained through the comparison of mortality rates, particularly if specific by age, sex and cause, for different geographical areas and different times. The data for such comparisons are generally readily available and involve large numbers of events.

The study of the geographical variations in mortality is relatively simple, requiring only consistency in data quality, the definitions used and the calculations. Consistent calculations of life expectancy for different countries are published by the World Health Organization (WHO). In Europe, the highest levels of life expectancy for men are observed in the Scandinavian countries and in Greece, showing that there is no simple geographical gradient in the mortality in Europe. Similarly, the study of time trends in mortality rates requires only consistency in data quality, definitions and calculations.

Variability over place and time can also be studied jointly. Figure XII.1 illustrates the stagnation/increase in mortality for middle-aged men in Europe. The age-adjusted mortality for males aged 40-64 years in the Netherlands increased in the 1950s and 1960s. In England and Wales, mortality declined throughout the period, whereas it remained nearly unchanged in the Federal Republic of Germany. Hungary experienced a decline in mortality in the 1950s and 1960s, but a steep increase in the 1970s.

Although the examination of very broad differentials such as these indicates undoubted differentials in risk factors to which the populations are exposed, it provides little or no indication of what the risk factors may be. For such purposes, the study of mortality rates specific for age, sex and cause of death, either over time or for much smaller geographical units than countries, is likely to be rewarding. Marked differential mortality from some particular cause between two adjacent areas may indicate some particular environmental or occupational risk factor present in one area but not in the other. Similarly, a marked change over time in a particular age-, sex- and cause-specific mortality rate may indicate important behavioural changes: trends in lung cancer deaths are an obvious example.

A study conducted at Chicago (United States of America) provides an example of the use of small-area information to examine socio-economic differentials in mortality. Five socio-economic groups within the city of Chicago were defined by assigning residents of each of the census tracts at Chicago to a socio-economic group on the basis of the median rent (1930 and 1940) or median family income (1950 and 1960) of the tract. The authors conclude that:
In general, despite the limitation of this approach which attributes to each individual residing in a given census tract the average characteristics for the tract as a whole, it has demonstrated that there are wide variations in mortality by socio-economic status within metropolitan areas. These studies, as those based on individual characteristics, reveal an inverse relationship between mortality and socio-economic status.\(^5\)

The possible pitfalls of examining small-area differentials are illustrated by a report\(^6\) from Finland on cancer incidence in relation to socio-economic characteristics of municipalities. This study found an increasing incidence of cervical cancer with increasing average income and percentage in the two highest social classes. This result is not in accordance with earlier studies, and the authors comment:

"The apparently conflicting results with regard to cancer of... the cervix uteri might be explained simply by the fact that, in the main, the population of the municipalities are heterogeneous, i.e., high-risk sections with a low standard of living are found even in modern industrialized environments. This also shows the problems involved in the interpretation of the results of studies like this."

The detection of differentials is only a first step in identifying the health problems involved. The second step in the process is to associate the observed differentials over place or time with behavioural or environmental differences or changes. Such studies are sometimes called "correlation studies". However, detailed individual-level studies have not always confirmed the apparent link between excess mortality and environmental or behavioural factors found in correlation studies. There are several examples of apparent relationships which it has not been possible to confirm at the individual level. One such example is a positive correlation at the national level between the incidence of cancer of the rectum, and to some extent also colon cancer, and beer consumption per capita; this relationship could not be confirmed in a follow-up study\(^8\) of brewery workers in Denmark having a beer consumption above the average.

B. CROSS-SECTIONAL STUDIES OF MORTALITY DIFFERENTIALS BY INDIVIDUAL CHARACTERISTICS

Mortality differentials between areas and time periods have proved very useful for suggesting the existence of specific health problems but, as has been demonstrated by the examples in the previous section, the possible confounding effects of uncontrolled factors on such differentials limits their value in establishing causal links. For such purposes, individual-level studies with more elaborate experimental designs are essential. This section and that which follows it describe and illustrate the principal approaches in current use. Their main characteristics are summarized in table XII.1.

**Conventional cross-sectional studies**

The cross-sectional study uses information on deaths and population for a particular point in time to examine mortality differentials by individual characteristics. The typical data sources are registered deaths to provide numerators and a population census to provide denominators. The decennial studies on occupational mortality for England and Wales provide typical examples. They are based on census data and data from death certificates. The population distribution by sex, age and occupation is known from the census and is used as the denominator in the calculations. The distribution by sex, age and occupation of persons dying in a period of three or five years around the census date is taken from the death certificates and is used as the numerator in the calculations. William Farr\(^9\) pioneered this type of study, first applied using the
1851 data; and the existence of a social class gradient in mortality has been documented for more than a century, as is shown in figure XII.2.

Some of the ambiguities inherent in geographical studies are avoided by the use of this approach, but problems remain. The definitions used for recording occupation on the census forms and on the death certificates have been co-ordinated so that both sets of data refer to last full-time occupation. However, discrepancies are discovered when death certificates are linked to individual census records. Among 607 men aged 15-64 years at the census in April 1971 who died in 1971, 233 were allocated to different occupational categories at the census and on the death certificate, and 62 were allocated to different occupational units within the same occupational order. More men were allocated to “inadequately described occupations” or “unoccupied” at the census than at death. A higher percentage than expected from pure chance of men “unoccupied” in the census were classified as labourers not elsewhere classified, miners or farmers on the death certificate, leading to an overestimation of mortality in these occupations.10

The importance of the documentation of persistent mortality inequalities in societies from studies such as these cannot be overestimated. The publications “represent an assessment of the current position. That coal miners die from pneumoconiosis is not a novel discovery—that they continue to die from this disease is a sad reflection on the inadequacy of preventive measures over the past 30 or so years.”11 However, due to the inconsistencies in the occupational classifications described above and the lack of information about possible exposure to risk factors earlier in life, such studies are less valuable for the identification of specific occupational or other risk factors. Different types of linking techniques have been devised in order to overcome this deficiency.

Retrospective unlinked studies

The problems arising from combining death certificate and census data in the analysis can partially be met by collecting data for deceased persons which are equivalent to the data collected for the risk population at the census. Such an approach will be used in the

---

TABLE XII.1. CHARACTERISTICS OF METHODS FOR THE STUDY OF OCCUPATIONAL MORTALITY DIFFERENTIALS

<table>
<thead>
<tr>
<th>Characteristics of method</th>
<th>Occupation known at different points in time</th>
<th>Mortality followed long after registration of occupation</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
<td>Linking</td>
<td>No</td>
<td>United Kingdom decennial survey</td>
</tr>
<tr>
<td>A. Cross-sectional studies</td>
<td>No</td>
<td>No</td>
<td>United States, 1980s</td>
</tr>
<tr>
<td>B. Follow-back surveys</td>
<td>No</td>
<td>No</td>
<td>United States, 1960s</td>
</tr>
<tr>
<td>C. Matched-record studies</td>
<td>Yes</td>
<td>No</td>
<td>United States, 1960s</td>
</tr>
<tr>
<td>D. Follow-up studies; cen</td>
<td>Yes</td>
<td>Yes</td>
<td>Scandinavian studies</td>
</tr>
<tr>
<td>D. Follow-up studies; cen</td>
<td>Yes</td>
<td>Yes</td>
<td>United Kingdom and Norway, samples</td>
</tr>
<tr>
<td>D. Follow-up studies; cen</td>
<td>Yes</td>
<td>Yes</td>
<td>Norway</td>
</tr>
<tr>
<td>D. Follow-up studies; cen</td>
<td>Yes</td>
<td>Yes</td>
<td>Occupational medicine, e.g., United Kingdom, polyvinylchloride workers</td>
</tr>
</tbody>
</table>

United States of America for the study of mortality differentials in the early 1980s. A mail questionnaire will be sent to relatives of a sample of 50,000-70,000 deceased persons. The occupational data for these deceased persons will be coded according to the 1980 census classifications, and the census data from a sample of 69,000 households will be used to provide denominators.\(^{12}\)

If this approach is extended to include information on exposure to risks earlier in life both for the deceased and for the household sample, the design is equivalent to a case-control study with population controls (see section D).

**Matched-record studies**

The approaches described above are not true individual-level studies, but rather attempts to classify events and exposure for relatively homogeneous population subgroups. As such, they benefit from large numbers of captured events, given the simple design, but suffer from the small number of factors that can be allowed for. The matched-record study, whereby the census records for a sample of persons dying in a particular period are linked to the death records, provides greater scope for disaggregation, although still on a cross-sectional basis. A study undertaken in the United States in 1960 provides an example.\(^{13}\) Deaths occurring in the United States during the period from May to August 1960 were manually matched to the individual census records from April 1960, using addresses. However, the census data on socio-economic characteristics were coded for only 25 per cent of the census population and mismatching therefore occurred at two stages in the mortality study: in the linking of death certificates to the general census records; and in the subsequent linking to socio-economic records for the deceased persons supposed to be included in the 25 per cent coded sample. Consequently, only 75 per cent of the death certificates were ultimately matched.

**C. FOLLOW-UP STUDIES**

Cross-sectional studies suffer from a limitation in that they provide little information about the earlier experience of the deceased. If death is regarded as the culmination of continuous exposure to various risks, such information is clearly of value. The follow-up study, whereby a population is defined by a census or baseline survey and then followed up for a period of time, provides such information and avoids biases introduced by, for example, disease-related changes in occupation prior to death. However, the costs of these advantages are long delays before results can be obtained and relatively small numbers of events to work with.

**Follow-up studies based on census data**

Censuses can be used to identify the base population to be followed up. The central technical problem in such a case is the matching of subsequent deaths to the initial study population. Personal identification numbers facilitate this process. In France, the Répertoire national d'identification des personnes—which is a manual register—has been used for a 17-year follow-up study of mortality in a sample of 500,000 men from the 1954 census.\(^{14}\)

The Central Population Registers in the Scandina-
vian countries have been used for mortality follow-up studies of the total 1970 census populations, three-year follow-up studies in Norway\textsuperscript{15} and five-year follow-up studies in Denmark\textsuperscript{16} and Finland.\textsuperscript{17}

The National Health Service Central Register for England and Wales—which is a manual register like that in France—has been used for a five-year follow-up study of mortality in a sample of 500,000 persons from the 1971 census.\textsuperscript{18}

In the Federal Republic of Germany, a one-year follow-up study of mortality is based on data collected from a sample of 181,000 households included in two successive labour force sample surveys, also called micro-censuses.\textsuperscript{19}

In each of the studies, mortality in the follow-up period has been analysed by occupation at the time of the census. The Scandinavian studies show that the 10 per cent of the men who were aged 20-64 years and economically inactive at the time of the census have a mortality rate that is 2.5 times the rate for the economically active. Within the economically active part of the population, male unskilled workers have a 40-50 per cent excess mortality in the younger age groups. This excess decreases with increasing age and disappears towards the pensionable age (see the diagrams from Denmark and Norway in figure XII.3).

One of the purposes of occupational mortality studies is to generate hypotheses about occupational risks. For this purpose, detailed tabulations by occupation and cause of death are necessary. The Danish study has, for instance, revealed an excess risk for lung cancer among butchers. This association is, however, not seen when all workers in the food and beverage industry are studied together.\textsuperscript{20} Detailed tabulations require large numbers, and although the total national populations are included in the Scandinavian studies, prolongation of the follow-up periods will obviously improve the analytical possibilities.

Another limitation in the census-based studies is that only the occupation at one point in time is known. Selection mechanisms influencing the composition of the labour force at the time of the census therefore have to be taken into account in the interpretation of the results. The problem is clearly illustrated by the mortality pattern among women: housewives have a 40 per cent excess mortality in comparison with economically active women, suggesting a strong health selection in the labour force rather than exposure to differential risks.

**Continuous longitudinal studies**

The follow-up approach can be modified to avoid some of the pitfalls of selection effects by recording not only deaths but important changes affecting the living through regular monitoring of the study population. Such continuous longitudinal studies require intensive and lengthy field-work, but they also offer control of a wide variety of factors. For example, selection out of the labour force of unhealthy persons can be accounted for to a certain extent in the analysis of occupational mortality by inclusion of data on previous employment.

In Norway, individual records from the 1960 and 1970 censuses have been linked together. Based on this linked file, the mortality differentials in the period 1970-1973 for occupational and social groups have been studied, including both persons who were economically active in 1970 and persons who were economically inactive in 1970 but were members of the labour force in 1960.\textsuperscript{21} Figure XII.4 shows the difference in the mortality between men in social group D (unskilled workers) and men in social group A (professionals) in Norway in 1970-1973. The graphs show that the recorded difference in mortality between the two social groups is enlarged when the persons who were economically active in 1960 but retired in 1970 are included in the respective social groups. This suggests that the narrowing of the social difference in mortality in the older age groups can be explained in part by early retirement of unhealthy manual workers.

However, the data from Norway also show that even when early pensioners are included in the analysis, the excess mortality of male unskilled workers is greater in the younger than in the older age groups. The same pattern is observed in England and Wales and in Denmark.

Several factors seem to contribute to this characteristic pattern in the mortality of male unskilled workers: first, an unfavourable health selection in the younger age groups; secondly, an excess mortality from violent deaths; thirdly, the early retirement of unhealthy manual workers; and lastly, almost no social difference for men over age 50 in mortality from circulatory diseases; at this age, half the number of deaths are due to circulatory diseases.\textsuperscript{22}

**Follow-up studies based on cohort data**

The difficulties in interpretation of cross-sectional mortality data can also be reduced by registration of persons from the moment when they enter a particular exposure group, such as an industrial or occupational category. This technique has been widely used in occupational medicine, especially in the United States of America and in the United Kingdom of Great Britain and Northern Ireland.

Supplementation of the data on date of entry to the cohort with data on date of exit provides the possibility for the study of mortality in relation to length of exposure. A well-known example is the British study\textsuperscript{23} of workers in the polyvinylchloride industry. Selection mechanisms have been illustrated by data for cohorts of such workers. Table XII.2 shows mortality for male polyvinylchloride workers by age and length of time since entry to the industry. The mortality among men aged 25-44 during the first five years after they entered the polyvinylchloride industry is only half the national rate, with a standardized mortality ratio of 45, but reaches the average level 15 years after entry to the
Figure XII.3. Mortality for males, by occupational groups, Norway, 1970-1973; and Denmark, 1970-1975

Key:
- Social group A
- Social group B
- Social group C
- Social group D
- Social group E

(Sstandard = all occupied men)
Selection out of the industry of unhealthy workers is illustrated by the mortality data for current and past employees, respectively, 15 years after both groups entered the industry. The recorded standardized mortality ratio for current employees is then 74 in contrast to 108 for past employees.

The selection into the industry of healthy workers and the selection out of the industry of unhealthy workers complicate the study of the effect on mortality of length of exposure to vinyl chloride monomers—the very aim of the study. A method for analysis suggested in the polyvinylchloride study in order to overcome these difficulties is to look at the mortality by length of exposure among workers who were all alive 15 years after entry and who had all left the industry by that time. By limiting the analysis to this group, both the effect of the selection into the industry and the effect of selection out of the industry are taken into account. The analysis showed a slight increase in mortality by

---

### Table XII.2. Standardized Mortality Ratios Among Male Polyvinylchloride Workers, by Age and Length of Time Since Entering the Industry, England and Wales

<table>
<thead>
<tr>
<th>Age Group</th>
<th>0-4</th>
<th>5-9</th>
<th>10-14</th>
<th>15+</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>25-44</td>
<td>45</td>
<td>73</td>
<td>57</td>
<td>104</td>
<td>62</td>
</tr>
<tr>
<td>45-54</td>
<td>37</td>
<td>79</td>
<td>70</td>
<td>74</td>
<td>65</td>
</tr>
<tr>
<td>55-64</td>
<td>32</td>
<td>51</td>
<td>99</td>
<td>92</td>
<td>79</td>
</tr>
<tr>
<td>65-74</td>
<td>23</td>
<td>30</td>
<td>76</td>
<td>112</td>
<td>96</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>37</td>
<td>63</td>
<td>75</td>
<td>94</td>
<td>75</td>
</tr>
</tbody>
</table>

length of exposure, the standardized mortality ratios being 104 for workers exposed 0-4 years, 116 for workers exposed 5-9 years and 117 for workers exposed 10-14 years.

The example given above clearly illustrates the problems in assessing the effects on health of exposure to potentially hazardous substances.

D. MORTALITY STUDIES BASED ON NUMERATORS RATHER THAN RATES

All the methods described above refer to situations where data both for the cases or the deceased and for the populations at risk are known, making possible the calculation of conventional rates. Proportional mortality studies and case-control studies are methods for bypassing the need for information about the populations at risk in situations where only data for cases or the deceased are known.

Proportional mortality studies

In a proportional mortality study, the proportion of deaths due to a certain cause in a specific occupational or other group is compared with the proportion of deaths due to the same cause in the total population. The method is valuable in the search for excesses in mortality from rare causes of death. However, it is important in the interpretation to remember that no conclusion concerning the absolute mortality level can be inferred from the data. The following example illustrates the problem.

A study made in Switzerland showed that the number of deaths due to ischaemic heart disease in relation to the number of deaths due to other natural causes among men aged 45-65 was highest for professionals and lowest for unskilled workers. An index equal to 100 for all men was 156 for the professionals and 82 for the unskilled workers. Mortality rates from England and Wales, however, show an excess in mortality from ischaemic heart disease for unskilled workers; and Scandinavian studies show almost equal rates for different social groups for males over age 50. The Swiss data do not necessarily indicate a risk distribution for ischaemic heart disease that is different from the distribution observed in other countries; the apparently conflicting results can be explained by the excess mortality for unskilled workers from other natural causes of death, which results in low proportional mortality figures for ischaemic heart disease even though their absolute rates are equal to or above the average.

Case-control studies

Case-control studies consist of matched samples of cases, i.e., persons with a specific disease; and of controls, i.e., persons from whom it is possible to estimate the proportion of persons in the population exposed to a specific risk. It is important to stress that although controls in case-control studies are often selected among patients with diseases other than those being studied, the purpose of the series of controls is to estimate the exposure rate in the population. Consequently, series of controls are selected for instance among patients with a disease supposedly unaffected by the exposure in question.

Case-control studies are especially suited for the study of risk factors in rare diseases. A case-control study in Sweden, for instance, revealed a relative risk of from five to six for soft-tissue sarcomas among men exposed to phenoxyacids and chlorophenols. Men selected from the population and death registers and matched with the cases on age and municipality constituted the series of controls. Persons who died from cancer were excluded from the controls because phenoxyacids and chlorophenols might also cause types of cancer other than soft-tissue sarcomas. Fifty-two cases diagnosed over an eight-year period were included in the study. Due to the rareness of this disease, it is very unlikely that this excess risk would have been discovered in a follow-up study.

However, spurious results can arise from case-control studies if strict rules for selection of controls are not followed. For example, data from the United States Third National Cancer Survey were analysed on a case-control basis, the fraction exposed to certain risk factors among persons with one type of cancer being compared with the fraction exposed to the same risk factors among persons with other types of cancers. This analysis "indicates an inverse association between breast cancer and smoking which illustrates an artifact caused by inclusion of exposure-biased patients in the control group (lung and other smoking-associated cancers)."

E. EVALUATION OF METHODS

Methods for the study of mortality differentials must be evaluated in relation to the purpose of the study. In relation to data on occupational mortality, it is decisive for the evaluation of the method whether the purpose is to sharpen the awareness of social inequalities in societies, to generate hypotheses about occupational risks or to lay the basis for specific preventive interventions.

The cross-sectional method used in the decennial supplements in the United Kingdom is, as illustrated, sufficiently precise for the detection of social-class differences in mortality. The method is generally considered to be inexpensive because it is based on already existing data. However, when computerized registrations with personal identification numbers are available, linking of individual census records and death certificate records is no doubt both less expensive and more precise than the separate coding of occupational information on death certificates. Unfortunately, the linking of individual records for total national populations is at the moment possible only in the Scandinavian countries, and the use of cross-sectional or
sample-based linking studies should therefore be recommended in other countries.

The generation of hypotheses about occupational risk has normally been stated as the main purpose of information systems on occupational mortality. The experience from the United Kingdom decennial supplement is, however, short of examples and—in addition to the documentation of the social-class differences—the supplements have mainly functioned as a monitoring system for already established or suspected associations between occupational exposures and specific diseases. The generation of hypotheses about occupational mortality and smoking habits among British doctors studied by Doll and Peto can be explained by neither confounding nor chance biases. Additionally, several other studies—e.g., of veterans and volunteers in the United States of America—have shown the same association. The association is strong, and the risk for smokers in relation to that for non-smokers increases with the number of cigarettes smoked. Lastly, when the British doctors stopped smoking, their lung cancer rate approached the rate for non-smokers.

However, few studies of the relationship between specific exposures and the development of human cancer fulfill all the criteria listed above. It is also important to remember that although true negative studies exist, epidemiological studies often have an inherent bias towards a negative result, primarily due to the selection mechanisms on the labour market as discussed above in section C.

Concerning decisions on preventive interventions, therefore, mortality studies have to be supplemented mainly with experimental data and with data for earlier detectable human reactions through mortality studies. Subsequently, mortality data are needed in order to monitor the effect of interventions.

As discussed in the foregoing sections, it is possible to overcome some of the shortcomings in studies of mortality differentials by the use of record linkage, registration of occupation at different points in time, prolongation of follow-up periods, etc. Therefore, it is strongly recommended that occupational data from censuses, unions, industries, etc., should be collected and stored in ways which would make it possible to meet these methodological demands.

### SUMMARY

Mortality studies in developed countries serve to identify high-risk groups and unfavourable trends. During the 1970s, important methodological innovations have come from studies of occupational risks, smoking, side-effects of drugs, etc. Advantages and shortcomings of the different methods must be evaluated in relation to the purpose of the study. Methodological demands of mortality studies in relation to decisions on preventive interventions have especially been discussed within the field of cancer. It is strongly recommended that occupational data should be collected and stored in ways which would make it possible to overcome shortcomings in the design of mortality studies.

### NOTES
