Methods of collecting morbidity statistics

Revised report to the Eurostat Task Force on ‘Health and health-related survey data’

Val Mason and Ann Bridgwood
Office for National Statistics
Room D1/19
1 Drummond Gate
London  SW1V 2QQ

Tel:  44 171 533 5384
Fax:  44 171 533 5300

13 August 1997
## Contents

<table>
<thead>
<tr>
<th>Summary</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Summary</strong> (including Figure 1)</td>
<td>iv</td>
</tr>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>2. Background</td>
<td>1</td>
</tr>
<tr>
<td>3. Description of the study</td>
<td>2</td>
</tr>
<tr>
<td>4. Incidence and prevalence measures</td>
<td>2</td>
</tr>
<tr>
<td>5. Sources of morbidity statistics in the EU, and their comparability</td>
<td>3</td>
</tr>
<tr>
<td>6. The main sources, and their uses</td>
<td>3</td>
</tr>
<tr>
<td>6.1 Health interview surveys</td>
<td>4</td>
</tr>
<tr>
<td>a) Cross-sectional population surveys</td>
<td>4</td>
</tr>
<tr>
<td>b) Panel/cohort surveys</td>
<td>8</td>
</tr>
<tr>
<td>6.2 Medical records / administrative statistics</td>
<td>8</td>
</tr>
<tr>
<td>a) Hospital records</td>
<td>8</td>
</tr>
<tr>
<td>b) Disease-specific registers</td>
<td>10</td>
</tr>
<tr>
<td>c) Death registers</td>
<td>13</td>
</tr>
<tr>
<td>d) General Practice (GP) records</td>
<td>13</td>
</tr>
<tr>
<td>e) Administrative notifications</td>
<td>15</td>
</tr>
<tr>
<td>7. Methodological approaches to obtaining EU-wide comparable morbidity statistics</td>
<td>16</td>
</tr>
<tr>
<td>7.1 The steps involved</td>
<td>16</td>
</tr>
<tr>
<td>7.2 Two examples of methodological approaches</td>
<td>16</td>
</tr>
<tr>
<td>a) Cancer statistics</td>
<td>16</td>
</tr>
<tr>
<td>b) Cardiovascular disorders</td>
<td>18</td>
</tr>
<tr>
<td>8. Further work on developing EU-wide comparable morbidity statistics</td>
<td>20</td>
</tr>
<tr>
<td>Appendices</td>
<td></td>
</tr>
<tr>
<td>A.1 Incidence and prevalence morbidity statistics, and their uses</td>
<td>21</td>
</tr>
<tr>
<td>A.2 Examples of prevalence estimates for diseases listed in WHO ‘Health For All’ Target 4.6 and their validity</td>
<td>25</td>
</tr>
<tr>
<td>References</td>
<td>27</td>
</tr>
</tbody>
</table>
Methods of collecting morbidity statistics

Summary

1. Under the EU’s proposed Action Programme on ‘Health Monitoring in the Field of Public Health’, reliable incidence and prevalence data will be needed to monitor disease-specific morbidity in the EU.

For some diseases, particularly chronic diseases or disorders such as some cardiovascular disorders, and for certain mental health problems, it can be useful to have both prevalence and incidence statistics. For other diseases, one type of statistic may be the priority. For example, incidence statistics may be most appropriate for cancer and congenital anomalies, but prevalence statistics for longer-term diseases or disorders, such as certain mental health, respiratory or musculo-skeletal problems.

2. The UK Office for National Statistics (ONS) was asked to prepare a paper considering the methodological aspects of integrating various sources of data on disease-specific morbidity. This paper provides an overview of different sources of such statistics, their uses, strengths and limitations, and their comparability across the EU. It concludes with some suggestions for further work.

3. The availability of sources varies between countries, as do the confidentiality rules governing access to them. The methods used to obtain particular incidence and prevalence data will tend to differ in different countries though, over time, they may converge.

4. The present study has looked at a number of sources of such data. It concludes that different types of data are suitable for different types of diseases. Not all sources are equally suitable to produce prevalence and incidence data.

5. The uses and limitations of three major sources of data (surveys, registers and GP data) are summarised in Figure 1.

6. Surveys are appropriate for estimating the prevalence of long-standing, chronic diseases with low fatality, such as some cardiovascular and respiratory conditions. They are less useful for investigating relatively rare or high fatality illnesses, such as some types of cancer. The availability of EU-wide comparable disease-specific survey information is limited.

7. Disease-specific registers are primarily a source of incidence data. They are only likely to be created for a limited range of diseases or disorders, for which there is pressure to have detailed morbidity analysis. These tend to be diseases associated with premature death, high mortality and/or long-term care needs.

In EU countries there are relatively few national registers. Local registers are sometimes used in making estimates of national incidence, but must cover sufficiently
large populations to give reliable estimates, and even so, these are clearly subject to the possibility of geographical bias.

8. Concern about data confidentiality has often limited access to GP records. Where this is overcome, it is possible to produce estimates of the incidence and prevalence of both acute and chronic diseases.

9. Other sources of data include hospital records and administrative notifications, e.g. for congenital anomalies and communicable diseases.

10. The steps for developing EU-wide comparable morbidity statistics can be summarised as follows:

• agree the aims and priorities for collecting particular morbidity data

• agree which diseases or disorders are to be covered, and whether incidence and/or prevalence measures are needed for each, and over what time span

• agree the precise concepts to be measured

• identify, or develop, standard definitions and/or coding frames (e.g. ICD codes)

• gather information about the available data from all EU countries, and about the meta-data (e.g. the sources, definitions and methods used)

• decide whether to use the data available, making it as comparable as possible, OR to develop new sources in some, or all, countries

• agree priorities for action.
### Figure 1: Summary of comparisons of three main sources of morbidity statistics

<table>
<thead>
<tr>
<th></th>
<th>Surveys</th>
<th>Registers</th>
<th>GP data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Coverage</strong></td>
<td>Cross-section sample, usually of people in private households</td>
<td>Total population with disease in a sample of hospitals or areas, or in the whole country</td>
<td>A sample of GPs, either national or local</td>
</tr>
<tr>
<td></td>
<td>Institution populations are more difficult to measure and integrate</td>
<td>If sub-national, best if the base population is stable and non-mobile</td>
<td>Samples subject to sampling and/or non-response bias or error</td>
</tr>
<tr>
<td></td>
<td>Samples subject to sampling and/or non-response bias or error</td>
<td>Cases may not be recorded if reporting is voluntary</td>
<td></td>
</tr>
<tr>
<td><strong>Type of estimate</strong></td>
<td>Prevalence (incidence possible)</td>
<td>Incidence (prevalence possible)</td>
<td>Prevalence and incidence</td>
</tr>
<tr>
<td><strong>Frequency</strong></td>
<td>Occasional measures (except in the case of continuous surveys)</td>
<td>Continuous monitoring</td>
<td>Continuous monitoring or occasional surveys</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Long-term commitment to register</td>
<td></td>
</tr>
<tr>
<td><strong>Good sources for</strong></td>
<td>Chronic diseases with low fatality</td>
<td>Chronic or acute diseases with high fatality, or need for long-term care</td>
<td>Chronic and acute diseases</td>
</tr>
<tr>
<td><strong>Can also collect information on</strong></td>
<td>Socio-demographic information</td>
<td>Some socio-demographic information</td>
<td>Prescribed medicines, treatment, hospital referrals, deaths</td>
</tr>
<tr>
<td></td>
<td>Household and living conditions, health-related behaviour, treatment, use of services</td>
<td>Have identifiers, so links with other sources (e.g. registers or censuses) are possible, subject to access</td>
<td>Some socio-demographic information may be available, although the quality may vary</td>
</tr>
<tr>
<td></td>
<td>Can investigate co-morbidity</td>
<td>Details of disease, treatment, and recovery</td>
<td>Information may be available on health-related behaviour such as smoking, though likely to be partial</td>
</tr>
<tr>
<td></td>
<td>Links with other sources (e.g. medical records, death registration) are possible, subject to consent</td>
<td></td>
<td>Can investigate co-morbidity</td>
</tr>
<tr>
<td><strong>Accuracy of</strong></td>
<td>Depends on respondents' knowledge, and on</td>
<td>Accurate where all (most) cases have</td>
<td>Doctors likely to record a diagnosis</td>
</tr>
<tr>
<td>identification of disease/disorder</td>
<td>Surveys</td>
<td>Registers</td>
<td>GP data</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td>whether reported diseases/ symptoms can be coded</td>
<td>been medically diagnosed</td>
<td>rather than symptoms, but strict diagnostic criteria are needed to ensure comparability</td>
</tr>
<tr>
<td></td>
<td>Not good for ‘sensitive’ diseases, such as AIDS, or cancer</td>
<td>Cases may not be identified, reported, or codeable</td>
<td>Diagnoses can be validated against data on prescribed medicines and hospitalisation</td>
</tr>
<tr>
<td></td>
<td>Can identify cases where health services have not been consulted</td>
<td>Strict coding criteria are required to ensure comparability</td>
<td>Some diseases, e.g. sexually transmitted diseases, or those not treated by the GP, may not be reported to GPs</td>
</tr>
<tr>
<td></td>
<td>Other data sources can be used to validate self-reported diagnoses, including health examination surveys</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Methods of collecting morbidity statistics

A discussion paper

A revised version of a draft paper considered by the Eurostat Task Force on ‘Health and health-related survey data’ on 23 April 1997.

1. Introduction

This paper describes methods of collecting disease-specific morbidity statistics. It includes methodologies for diseases such as cancer and cardiovascular disorders, congenital anomalies (such as Downs Syndrome) and communicable diseases (such as malaria and measles). It does not cover general morbidity measures, such as those for ‘general health status’, functional status or disability.

The main statistics considered are prevalence and incidence. Incidence-based statistics such as recovery and survival rates are mentioned only briefly. The paper does not discuss the burden or consequences of disease, or cover measures of disease-free life expectancy and quality of life.

The paper provides an overview of different sources of disease-specific statistics, their uses, strengths and limitations, and their comparability across the EU. It focuses mainly on data from health interview surveys, disease-specific registers and general practitioner (GP) records.

To illustrate how statistics from different sources can be used either separately or in combination for different types of disease, Section 7 describes various sources for two diseases - cancer and cardiovascular disorders.

2. Background

EU countries are asked to monitor progress towards Health For All (HFA) targets on morbidity as part of the monitoring programme for the European Region of the World Health Organisation (WHO) 1993-94. HFA Targets 4 and 5 refer specifically to the prevalence and incidence of particular diseases (WHO 1993). Data are collected by WHO’s regional office for Europe in Copenhagen, and form the basis of WHO monitoring reports (WHO, 1994a & 1994b).

Morbidity indicators are also to be developed under the EU’s proposed Action Programme on ‘Health Monitoring in the Field of Public Health’ (DG V, 1993; DG V and Eurostat, 1996). If the Council and the European Parliament adopt in June 1997 the proposed Decision on this Action Programme, work on the five-year programme will start in the second half of 1997. At the time of writing, the details of this proposed programme are still to be formulated, but it is clear that there will be a need to monitor the incidence and/or prevalence of specific diseases in EU countries. Eurostat have therefore started to investigate some of the methodological issues involved.
The Task Force on ‘Health and health-related survey data’ is one of three Task Forces set up in 1996 by the Working Group on Public Health Statistics (DG V and Eurostat, 1997), with the general aim of improving the comparability of survey data and morbidity statistics in EU Member States. It first met in December 1996 and, among other topics, discussed a draft paper ‘Towards a Harmonised Morbidity Statistics Indicators System’ (Egidi, 1996), which explored the possibilities for integrating survey and registration data to provide morbidity statistics.

The present study was commissioned to consider the methodological aspects of integrating various sources of data on disease-specific morbidity. The study was intended to build on existing experience, primarily in EU countries, and to focus on methods rather than on particular diseases.

The study will set the basis for commissioning a wider-ranging review of morbidity statistics. The aim is to develop appropriate methodologies for collecting a range of disease-specific morbidity statistics within EU countries and for increasing the comparability of data between countries.

3. Description of the study

The project was carried out by Social Survey Division of the Office for National Statistics (ONS) in the UK. As well as consulting widely among staff in ONS responsible for national surveys and registration data, contact was made with a number of experts in Italy, Sweden, Denmark, the Netherlands and the UK.

The paper draws on discussions held with and information received from these contacts, and on a range of publications. However, no attempt has been made to conduct a comprehensive literature review, and further work will be needed to provide a complete picture of existing methodologies and of the possibilities for developing comparable morbidity statistics in the EU.

A draft version of this paper was discussed at the April 1997 Task Force meeting, and at a meeting of experts which was held in London in May. This revised version of the paper incorporates comments made at both meetings.

4. Incidence and prevalence measures

There are two basic types of morbidity statistic - measures of incidence and of prevalence. Incidence is a measure of the number of cases arising in a population in a period, and prevalence is a measure of the number of cases existing at a time or in a period (Ruwaaard et al, 1994). For example, incidence might be expressed as the number of new cases of a disease (or disorder) per 1,000 population in a year, and prevalence as the proportion of a population with the disease at any time in a year.

Incidence may refer either to the first onset of a disease (i.e. new cases) or to all episodes. Incidence measures provide the basis for statistics such as disease-specific mortality rates, and recovery and survival rates.
Because the incidence and prevalence of many diseases differ with sex and age, data are generally presented separately for different sex and age-groups. If, however, the aim is compare estimates for different populations, a summary figure can be calculated, which is standardised for sex and age, so that differences between the sex and age distributions are taken into account.

Details of the uses of these statistics are described in Appendix 1.

5. Sources of morbidity statistics in the EU, and their comparability

There are a number of possible sources of disease-specific morbidity statistics, although the extent to which these sources exist in different EU countries varies greatly.

There are also differences between countries in the confidentiality rules governing access to data, and the extent to which data from different sources can be linked (INED, 1995). It follows that the methods used in different countries for obtaining particular incidence and prevalence data will tend to differ, though over time, methods may converge. In the meantime, it will be necessary to accept that sources and methods may differ; the possible effects of those differences on the data must be understood, and should be taken into account in comparisons between countries.

A priority in developing EU-wide comparable statistics is to agree the concepts to be measured. Standard definitions are needed for the diseases (e.g. ICD codes) and for the other factors under investigation.

6. The main sources and their uses

The main sources of disease-specific incidence and prevalence data across a range of EU countries are as follows:

- Health interview surveys
  - Cross-sectional population surveys
  - Panel/cohort surveys

- Medical records / administrative statistics
  - Hospital records
  - Disease-specific registers
  - Death registers
  - General Practice (GP) records
  - Administrative notifications

Figure 1 gives a summary comparison of the main sources: surveys, registers and GP records.

6.1 Health interview surveys
Survey data tend to be used for producing estimates of prevalence rather than of incidence; Egidi (1996) argues that the sample size is usually too low for calculations of incidence to be made. Appendix 2 discusses the use of survey data for producing prevalence estimates of the diseases listed in WHO HFA Target 4.6.

a) Cross-sectional population surveys

Disease-specific morbidity data can be collected by different types of surveys, including health interview surveys (HIS), health and lifestyle surveys, multi-purpose surveys which include a health module, and surveys focusing on particular topics such as psychiatric morbidity, or disability.

(i) Alternative types of questions, and their effects

There are several possible approaches to collecting disease-specific information:

- Informants can be asked if they have any illness or disability. Those who do are then asked for details, using an open question, such as ‘What is the matter with you?’. This approach is used by the Health Survey for England, the UK General Household Survey and the Danish Health and Morbidity Survey 1994. Information from the open question is then coded, using codes which broadly correspond to ICD categories.

- When probing for details of morbidity, interviewers may ask about illness(es) or about symptoms. On the whole, respondents find it easier to report symptoms than underlying causes, partly because it is symptoms which affect people’s day-to-day lives (Foster et al, 1990), but it is not always easy to convert these descriptions to ICD codes (Ruwaard et al, 1994). Some symptoms, such as breathlessness, may be associated with more than one disease, and diseases or disorders can be formulated in non-medical terms (for example, describing a neoplasm as a ‘growth’).

- Informants can be presented with a checklist of diseases, and asked if they have had any of them. Examples include the UK Health and Lifestyles Surveys, the Welsh Health Survey, and the Netherlands Health Interview Survey (which asked about 24 diseases and disorders).

- Detailed questions can be asked about specific diseases, such as cardiovascular disease (Health Survey for England, using questions based on the Rose-Angina questionnaire), asthma and allergies (Danish Health and Morbidity Survey [DICE, 1996]) or psychiatric morbidity (See World Psychiatric Association, 1993 for examples).

Questions may or may not refer to a specific time period. Core questions are sometimes supplemented with further questions on Activities of Daily Living, which allow estimates of functional status or healthy life expectancy to be made.

(ii) Advantages of survey data
An advantage of surveys is that they can collect a wide range of information about respondents. They can include questions on, for example, respondents’ characteristics, and their use of health services. Data on health-related behaviour such as smoking, alcohol consumption, levels of physical activity and nutrition, are also often available. This enables researchers to identify associations between risk factors and specific diseases.

Because they often collect data about a range of diseases, surveys also offer the opportunity for exploring co-morbidity. The UK Surveys of Psychiatric Morbidity examined the co-morbidity both of different neurotic disorders, and of physical and neurotic conditions. A hierarchy of neurotic disorders was used to establish the primary disorder (Meltzer et al, 1995b).

The possibility of integrating survey data with other sources of data exists in several Member States. In Sweden and Denmark, for example, if respondents’ Personal Identification Number (PIN) is collected during a survey, the data can be linked to national registers, subject to informants’ consent and approval by the Data Registrar. In the UK, respondents can (with consent) be ‘flagged’ on the National Health Service Computerised Register - this enables survey data to be linked with information from cancer and death registrations.

(iii) The validity of survey data

Estimates of prevalence are influenced by a number of factors, including:

- **The survey context.** There is evidence that, even when identical question wording is used, higher prevalence estimates are produced by Health Interview Surveys than by multi-purpose surveys such as the UK General Household Survey (Bennett et al, 1995).

- **Question wording.** For example, surveys which attempt to measure both limiting and non-limiting chronic illness with one question tend to produce lower overall estimates of prevalence than those which ask two separate questions.

  Asking whether respondents ‘have’ a long-standing illness produces higher estimates than asking whether they ‘suffer’ from an illness; some people may answer ‘no’ to the latter on the grounds that they are not actually suffering (Goddard, 1990).

  Asking whether an illness limits activities compared with ‘people of your age’ produces lower estimates than asking respondents whether it limits them ‘in any way’; it is believed that elderly people in particular would say no because they consider that most of their contemporaries are as limited in their activities as they are (General Household Survey 1972).

- **Methodology.** Using a checklist has the advantage of providing all informants with a common frame of reference. It could, however, result in overestimates
of prevalence if informants who are not sure whether they have a condition include themselves (Goddard, 1990).

- **Respondents’ awareness** of their condition or understanding of the question wording. Those who have not been diagnosed or who do not recognise terms on a checklist may not report a disease which is present.

- **Type of disease.** Checks on validity have shown that the highest rates of agreement between survey data and doctors’ diagnoses can be expected for conditions which require ongoing treatment, have commonly-recognised names and are salient to informants because they cause discomfort or worry (Bennett et al, 1995).

- The **social characteristics** of respondents. There is some evidence that informants in manual social classes, particularly men, are likely to say they have a named disease only if it is actually troublesome; this is particularly true for mental disorders (Bennett et al, 1995; Blaxter, 1990).

Methods of assessing the validity of interview data on specific diseases include asking respondents who have reported a condition whether it has been diagnosed by a doctor. Survey data can also be compared with information from sources external to the survey, such as, with respondents’ permission, information from their GP. Discrepancies between the two sources do not necessarily indicate that self-reported data are inaccurate; informants may not have brought a condition to the attention of a doctor, medical records could be inaccurate, doctors may not have informed patients of their diagnosis, and lay descriptions may differ from those given by doctors (Blaxter, 1990).

Self-reported data can be compared with other data collected by the survey such as health examination data, which provide a potentially powerful source of ‘objective’ data; the results of clinical examination or data on prescribed or other medication. Where there are several sources of data, such as self-reported and health examination data and data on prescribed medicines, estimates of prevalence can be made with a higher degree of confidence. Appendix 2 gives some examples of the validity of survey estimates for some of the diseases in WHO Targets 4 and 5.

**(iv) Limitations of survey data**

Household-based surveys are liable to under-estimate the prevalence of more serious diseases, as sufferers are more likely to be in hospital or unavailable for interview (Blaxter, 1987; Ruwaard et al, 1994). When surveys do cover both the household and the institutional population, as was the case with the UK surveys of Disability (Martin et al, 1988) and of Psychiatric Morbidity (Meltzer, 1996a), separate estimates are usually made for the two populations; no overall estimate of prevalence is made.

Surveys are appropriate for producing prevalence estimates of long-standing, chronic diseases with low fatality, such as some cardiovascular and respiratory conditions. They are less useful for investigating relatively rare, acute, high fatality or short-
duration illnesses such as some cancers, because surveys rarely yield sufficient cases for analysis. Estimates will be biased by high mortality and non-response because of poor health and hospital stays. In addition, people with conditions such as cancer may be unaware of the true nature of their condition (Egidi, 1996) or unwilling to talk about it. ‘Neoplasms’ have been identified as particularly difficult for surveys to code to ICD classifications (White et al, 1993).

(v) The comparability of survey estimates in the EU

How feasible is it to make international comparisons of survey data on disease-specific morbidity? Egidi (1996) notes that methodological and organisational survey procedures such as sample size and structure, questionnaire design and data collection instruments, limit comparability between countries. A key to making valid comparisons is the distribution of meta-data, such as survey definitions, as well as actual data.

The European Division of WHO, together with the Central Bureau of Statistics (CBS) Netherlands, started the Health Interview Survey project in 1987, with the aim of developing common methods and instruments for health interview surveys. These are used as a framework for collecting and adjusting already-existing data from different countries and, in the longer term, WHO is promoting their adoption by countries (Nanda, 1997). The returns to a ‘survey of surveys’ carried out by WHO Europe in 1995-6 indicate that a number of countries collect data on chronic physical and mental conditions, but as there is no recommended instrument, data from different national surveys are not currently readily comparable (Cambois and Robine, 1997). Eurostat is also currently compiling an inventory of items covered by national surveys, which is complementary to the ‘survey of surveys’; a report on this project was discussed at the April meeting of the Task Force (Eurostat, 1997).

Some survey data relating to health indicators are available at a Community level; the European Community Household Panel (ECHP) survey, for example, includes four health indicators, but they do not provide information about specific diseases. A supplement on health (though again, not disease-specific) has been added to the Eurobarometer. This has the advantage of speed and flexibility, but is limited by the relatively small sample size (DG V, 1997).

b) Panel or cohort surveys

The main advantage that panel or cohort surveys have over one-off surveys is that they re-interview their respondents at specified intervals. They can, therefore, track any changes in their morbidity over time and can relate these to other changes in lifestyle or circumstances. However, such surveys tend to be methodologically complex and expensive. They also tend to suffer from attrition over time, with an attendant risk of non-response bias, and could produce a ‘panel effect, whereby the process of being repeatedly interviewed can influence either behaviour or reporting.

A number of EU countries have long-running cohort studies, but where the survey designs and times between interviews differ, the data are not likely to be comparable
(Egidi, 1996). To obtain comparable data, surveys with similar methodologies would have to be carried out in each country. One possibility is to consider adding questions asking about specific diseases (either by including a checklist or by following the currently-asked open question about chronic illness with a further question asking for details of the conditions) to a panel survey such as the ECHP Survey. The April 1997 meeting of the Task Force discussed the possibility of including some questions on Member States’ Labour Force Surveys.

6.2 Medical records / administrative statistics

Medical records and administrative statistics provide information about the numbers of people with certain diseases/disorders. To calculate incidence and prevalence, all these sources must be supplemented with information about the base population, e.g. the numbers of people in a range of age and sex groups resident in a specified geographical area. These estimates may be derived, and updated, from sources such as censuses or population registers.

a) Hospital records

(i) Hospital admission or discharge records

Hospital admission or discharge records can contribute to disease-specific incidence data, but only to a limited extent.

They have the advantage over many survey sources of identifying diseases on the basis of medical diagnosis, (though medical notes do not always provide complete data recorded in a consistent way).

A major limitation is that they can only be the main source of data for diseases where hospitalisation is likely to occur. Other sources are needed to identify any non-hospitalised cases.

A further limitation is that hospital records can only measure incidence of new cases if people’s first admissions to hospital for the disease are reliably identified. This is not the case, for example, in England’s Hospital Episode Statistics, which are episode-based, not person-based (Department of Health, 1996). However, the use of personal identifiers to link records can overcome this as in, for example, Sweden.

Hospital admission or discharge records are used mainly in conjunction with other sources to help obtain complete coverage of incidence, particularly in compiling disease-specific registers. Examples include national registers for cancer in several EU countries, for myocardial infarction in Sweden, and for mental disorders in Italy, as well as local registers. (See also Sections 6.2b and 7.)

(ii) Hospital records of births (and stillbirths)

Hospital records of births (and stillbirths) provide almost complete coverage of births in most EU countries. If supplemented with information about home births, they are,
therefore, a potentially good source of information about the incidence of congenital anomalies. However, it is difficult, in practice, to obtain complete coverage of cases. Some cases may not be diagnosed initially, others not recorded, and others not reported to the registry. Standardised definitions and methodologies are needed, and comprehensive checks must be made on data.

The extent to which hospital birth records are used varies between countries. In some, information is collected only from a limited sample of hospitals, while in others, e.g. Denmark, hospital records are used to compile mandatory, national registers, which are linked to other registers to add epidemiological information (WHO, 1991 and EUROCAT, 1991).

In the UK, national incidence statistics for congenital anomalies have traditionally been obtained from local health areas. Their main source of data have been hospital ‘Notification of Birth’ records, made within ten days of births. However, this method has identified only about half of congenital anomalies. To improve the quality of the data, the ten-day limit has been removed and links are to be made with local registers of cases. These registers provide more complete data than Notification of Birth records, as they are used at local level (for research and/or to track cases and plan resources), and so have established systems for collecting ‘late’ information about anomalies not identified at birth, and for updating data.

b) Disease-specific registers

Disease-specific registers are primarily a source of incidence data, though methods exist for estimating prevalence from them in the absence of other sources (Egidi, 1996, and Verdecchia et al, 1989). (See also Section 7.2a.)

(i) Limitations on the range of diseases/disorders

Registers are only likely to be created for a limited range of diseases or disorders. They are often a labour-intensive and costly source of statistics, involving continuous record-keeping over a number of years. They are generally set up only for diseases for which there is pressure to have detailed morbidity analysis, (of the types outlined in Appendix 1); for example, diseases associated with premature death, high mortality and/or long-term care needs.

---

1 Although disease-specific registers are costly to set up and maintain, they can be a very cost-effective source of a wide range of data.

2 Registers of survey respondents can be set up in, for example, Sweden for the purposes of a particular study, but are dismantled once the study is completed. This section of the paper is not concerned with this type of register.

9
In addition, because cases are generally identified through contact with health care services, registers do not provide complete data for diseases where a significantly large proportion of sufferers are not known to those services.

(ii) Methods of compiling registers and data confidentiality

Disease-specific registers are normally based on information collected for health care records; for example, by hospitals or GPs. They generally aim to identify all new cases of a disease, either as they occur or retrospectively, and to record key information about them. This can involve obtaining and matching information from a wide range of sources. Registers are complex to compile, and can require a good deal of effort to ensure complete and accurate coverage. For example, in registers where complete coverage is required for research purposes, some cases can only be identified from ‘cause of death’ information in death records, and so involve retrospective tracking to obtain incidence data.

Registers, by definition, have a means of identifying individuals, either with names and addresses or with personal identification numbers (PINs). This permits the linking of data about an individual from different sources and over periods of time. For example, data may be linked with one or more of the following - medical records (held by hospitals, GPs and others, and giving initial diagnosis and/or subsequent treatment and recovery), population registers, census data, survey data, and death registers. However, the extent to which such links can be made is determined by data protection laws, which vary across Europe. For example, in some southern European countries, death registration records are confidential and are not directly available to registries, whereas in other countries permission is given for links between several sources.

(iii) Coverage of cases

In principle, registers include all cases of a disease which come into contact with the organisations providing data to the registry. In practice, ensuring complete coverage is difficult. Some cases may not be diagnosed initially, others not recorded fully, and others not reported to the registry. Many registers rely on voluntary reporting of cases and, in Germany for example, individuals’ consent is required before their case is added to a register. Coverage will tend to be improved if made mandatory (Esteve, 1993).

(iv) Accuracy of data

Registers have the advantage over many survey sources of identifying diseases on the basis of medical diagnosis. Although medical notes do not always provide complete data recorded in a consistent way, cross-checks with different sources of cases can improve the accuracy of a register.

(v) Background information about individuals

As well as information about the disease, registers are normally able to collect key information about individuals, such as their sex, age, area of residence, occupation,
and possibly treatment regimes. The amount and quality of such information is sometimes improved by links to other sources of data, such as birth or death registers, population registers or census records, or to summary census data for the individual’s postcode area.

(vi) National and local registers

Registers may be national or local depending on the reasons for setting them up, and on the resources available. National registers are used to monitor national patterns and trends in incidence and survival rates, particularly for diseases where there is a demand for continuous monitoring of short-term changes in patterns and trends across the different areas of a country.

In EU countries there are relatively few national registers. The diseases most often covered are congenital anomalies and cancer. There are also national registers in Sweden for myocardial infarction, and in Italy for people admitted to psychiatric hospitals. There is also some interest in developing registers for renal diseases. The possibilities for developing national registers will vary between countries, depending on priorities and on factors such as population size and data protection considerations.

Local registers are often set up to focus on specific aetiological issues, or for the assessment of differing treatment regimes. In addition to the diseases covered in the national registers mentioned above, local registers exist to cover diseases such as diabetes, asthma and allergies. Compared with national registers, they have the advantage of being on a smaller scale, so are less expensive and have potentially greater control over data quality. However, they require a relatively stable, non-mobile population if they are to avoid having to trace individuals moving in or out of the region, and having to take account of changes in the base population.

Local registers are sometimes used in making estimates of national incidence, but must cover sufficiently large populations to give reliable estimates, and even so, they are clearly subject to the possibility of geographical bias (Esteve, 1993).

(vii) Alternative approaches to obtaining comparable register data

There are two main approaches to obtaining comparable national data from disease-specific registers. These are:

- Using the data which does exist in the different countries, taking account as far as possible of the disparate definitions and methods, and attempting to derive comparable estimates. This method builds on national systems already in place and should ideally draw on the expertise existing in each country, though for those countries which lack national systems, estimates must be derived in other ways.

Examples of this method are the work carried out by The International Agency for Research on Cancer (IARC) on cancer incidence in Europe (Esteve, 1993),
and that on congenital anomalies carried out by the International Clearing House for Birth Defects Monitoring System (WHO, 1991).

- Setting up a project covering all countries to collect specified data using standard definitions and methods to ensure that the data collected are as comparable as possible. Because of the costs involved, however, such projects will generally be limited to collecting information at sub-national level, from only a limited number of local areas. While this may be acceptable given the main aims of the studies, it is less satisfactory for producing national estimates, if the data are not from a representative sample of areas, and so are subject to sample bias. Even with these limitations, the results may point to possible differences between countries, and is information about local populations and environmental factors is available, some assumptions about national differences can be made.

Examples are the MONICA cardiovascular disease research project (WHO, 1989), and the work of the EUROCAT network of regional registries for congenital anomalies (EUROCAT, 1991).

c) Death registers

Although mainly a source of mortality statistics, the ‘cause of death’ information in death registers can, in the absence of other sources, be used to estimate the incidence of certain diseases. This is only the case for diseases associated with low recovery rates and high mortality, and where estimates of survival times can be made. For example, this has been done for cancer incidence for certain countries (Esteve, 1993: Parkin, 1993). (See also Section 7.2a (i).)

d) General Practice (GP) records

Concern about data confidentiality has often limited access to GP records. Where this is overcome, it is possible to produce estimates of the incidence and prevalence of both acute and chronic diseases. GP records include information on consultations, diagnoses, prescriptions, treatment and test results. They may also include records of hospitalisation and referrals and notifications of death.

There are two main approaches to collecting disease-specific data from GPs: GPs can contribute information to a computer database, or participate in surveys. Both approaches are used in England and Wales, and in the Netherlands.

In England and Wales, the General Practice Research Database (GPRD) was established in 1987. It covers over 500 practices, which use a common computer system, and almost three and a half million patients (ONS, 1996a). The Dutch Continuous Morbidity Registration (CMR) draws on records from general practitioners in the Nijmegen region (Ruwaard et al, 1994). Both countries also carry
out surveys of GPs; the UK Morbidity Survey of GPs and the Dutch National Survey on Morbidity in General-Practice (NIVEL).

GP data have been used to produce estimates of the prevalence and incidence of specific diseases. (Appendix 2 gives examples of prevalence estimates produced for diseases listed in WHO HFA Target 4.6.) In the Netherlands, the national study (NIVEL) only covered a period of three months. Therefore, while national estimates for annual incidence were derived (by multiplying the results by four), no national figures for annual prevalence could be derived (Ruwaard et al, 1994). The EuroSentinel project, which involved a number of EU countries, was set up to produce international estimates from GP data of the incidence of measles, mumps, HIV-antibody tests and influenza-like illness (van Casteren, 1993).

(i) Advantages of GP data

GP records, like registers, are more likely than surveys to provide a medical diagnosis. However, strict diagnostic criteria need to be followed by GPs if data are to be comparable. Data on medication and hospitalisation can be used to improve the accuracy of data.

A longitudinal database can be created from GP records, effectively forming a GP-based register. This enables changes in morbidity to be monitored. Where GPs are required to notify the database of patient deaths (as in England and Wales), information can be used to calculate survival rates.

Information extracted from GP records can cover a range of diseases, so can be used to explore co-morbidity. This gives it an advantage over register information, which is usually limited to one type of disease. Co-morbidity is being investigated in this way in the Netherlands. Although the work is still in its infancy, the data suggest that the degree of co-morbidity varies with diagnostic group; it is relatively high for some of the diseases covered by WHO HFA Target 4.6, including cancer, musculo-skeletal and pulmonary diseases. There is also some clustering between pulmonary diseases, coronary heart diseases, rheumatic diseases and diabetes, (Ruwaard et al, 1994).

GP records may include some socio-demographic data, although the quality of this is likely to be variable. It will not necessarily be recorded in a standard form by all participating GPs, particularly in different EU countries. Some information on health-related behaviour, such as smoking, alcohol consumption and diet, and on obesity may also be available.

(ii) Limitations of GP data

A limitation of GP data is that, depending on the sample of GPs included, the coverage may be unrepresentative of either patients or diseases. The Dutch CMR, for example, is a regional database and it is not clear whether this region is representative of the Netherlands as a whole. The NIVEL study did not cover people who had not seen their GPs.
Although the level of registrations with a GP may be high (98-99% in the UK, for example), a UK study of psychiatric morbidity among the homeless found that, among those using night and day shelters, three fifths to two thirds were registered with a doctor, but that this was not necessarily the doctor they currently had access to (Gill et al, 1996).

Some minor diseases, and even major ones (e.g. those treated only in hospital) may not be reported to GPs. Specific conditions, such as sexually-transmitted diseases, may be reported to agencies other than GPs.

The quality of data recording may also be variable; although it can be improved by providing participants with recording guidelines (Hollowell, 1997). Some checks are required on data quality. The experience of EuroSentinel showed that diagnoses were not always reliable (van Casteren, 1993); differences in diagnostic criteria were particularly evident for influenza-like illness.

A major limitation of using GP data to produce international comparisons is that not all EU countries have a similar system of general practitioners, who in addition to keeping their own records, receive data about their patients from external sources, such as hospitals and family planning clinics. Even where GP systems are similar, records would have to be kept in a consistent way, and account taken of national differences in diagnoses and prescribing practices is EU-wide estimates are to be made. In the EuroSentinel project, the term ‘GP’ was interpreted in different ways (van Casteren, 1993).

e) Administrative notifications

Administrative notification systems exist to provide up-to-date Public Health information about short-term trends and epidemics of communicable diseases. In the UK, for example, doctors are required to inform local government officers each week of all cases of about thirty infectious diseases, e.g. malaria, measles, and food-poisoning. This information is then collected centrally, without names and addresses, to produce national incidence statistics (ONS, 1996c). In a similar way, the incidence of sexually-transmitted diseases (and of HIV/AIDS) is obtained from voluntary and confidential reports from doctors and others identifying cases, e.g. at clinics or laboratories.

Similar systems exist in other EU countries. The data available from different countries on food-poisoning is reported as part of the WHO Surveillance Programme for the Control of Foodborne Infections and Intoxifications in Europe (Institute of Veterinary Medicine, 1992).
7. Methodological approaches to obtaining EU-wide comparable morbidity statistics

7.1 The steps involved

The steps for developing EU-wide comparable morbidity statistics can be summarised as follows:

- Agree the aims and priorities for collecting particular morbidity data
- Agree which diseases/disorders are to be covered, and whether incidence and/or prevalence measures are needed for each, and over what time span (See Ruwaard et al, 1994)
- Agree the precise concepts to be measured
- Identify, or develop, standard definitions and/or coding frames (e.g. ICD codes)
- Gather information about the available data from all EU countries, and about the meta-data (e.g. the sources, definitions and methods used)
- Decide whether to use the data available, making it as comparable as possible, OR to develop new sources in some, or all, countries
- Agree priorities for action.

7.2 Two examples of methodological approaches

For some diseases/disorders, it is useful to have both incidence and prevalence statistics, while for others, one type of measure is the most useful. Generally for short-term acute illnesses, estimates of incidence and their related statistics are more useful than prevalence measures (See also Appendix 1).

To illustrate how statistics from different sources can be used either separately or in combination for different types of disease, sources for two diseases - cancer and cardiovascular disease - are described below.

a) Cancer statistics

For cancer, incidence statistics are the main priority though prevalence measures are also of interest.

(i) Incidence
Cancer is a disease associated with high treatment costs and, for some types of cancer, short survival times and high mortality. There is pressure for research into the different types of cancer (e.g. lung, breast, skin) to establish causes, risk factors, the effectiveness of treatment regimes, and recovery and survival rates. This requires incidence-based statistics and links to death registers.

The numbers of cases, particularly of rarer types of cancers, are relatively small so the identification of all new cases in a region, or country, is often necessary to yield sufficient numbers for analysis. These numbers are not generally too large to over-burden reporting and recording systems, as might be the case for more common diseases.

Cancer registers are the main source of incidence data. GP records might increasingly be used, though the numbers identified may be small. Population surveys will generally yield insufficient numbers, except of the most common types of cancer. More importantly, they cannot identify fatal cases, and will be subject to sampling and reporting bias.

Cancer registries exist in most EU countries, but they are national in some Member States, e.g. the Netherlands, the Nordic countries, and the UK countries. However, progress is being made in producing comparable estimates of cancer incidence across Europe, and world-wide. In 1990, a European Network of Cancer Registries was launched within the framework of the EC ‘Europe against Cancer’ programme. Its work is co-ordinated by a steering committee and is centred at the International Agency for Research on Cancer, IARC, in Lyon. Its aim is ‘to improve comparability of cancer registry data and extend cancer registration in the EC’. It reports on the patterns of incidence and mortality by sex, age and type of cancer (Esteve, 1993).

To obtain estimates of national incidence in those EU countries with no national registries, IARC have used estimation models based on mortality data from death registers. These use information about the ratio of incidence to mortality, where known, from groups of local registries. They take into account age, sex and broad geographical region of Europe (north, south and west). Updated estimates are based on changes in mortality in each country (Esteve, 1993). These estimates, though a start, are far less satisfactory than obtaining actual national estimates. For example, they do not reflect any differences in incidence and mortality ratios between the different regions within countries or between the different countries within the broad regions. Also, they are subject to variations in the reporting and coding of ‘cause of death’ in death registers.

(ii) Prevalence

The prevalence of cancer tends to be relatively low, reflecting low survival rates for some types of cancer. For example, in England and Wales, it has been estimated that only about 1 - 1.5% of adults living in private households ‘have’ cancer at any one time, and many of these will have been cured and are not a burden on the Health
Service. For this reason, measures of prevalence, although of interest, are less important than measures of incidence.

Estimates of the prevalence of cancer can be obtained in a number of ways, though each method presents difficulties. Survey estimates are subject to error for a number of reasons, such as high mortality, non-response because of poor health and hospital stays, and respondents not knowing about their cancer or not wanting to talk about it. GP records might increasingly be used, though the numbers identified may be small.

Cancer registry data can be used in two ways to estimate prevalence, as follows.

- Identify actual cases in registers and, through links with death registers, track when the people die. Counts of the numbers still alive with cancer can then be made at any particular time.
- Produce estimates by modelling, based on incidence and recovery/survival rates (Egidi, 1996; Verdecchia, 1989). A proposal has been put forward to estimate cancer prevalence in this way in each EU country (Capocaccia, 1996). One difficulty envisaged by the researchers is that registers seldom started early enough (e.g. pre-1970s) to identify all current cancer survivors. This problem might be avoided if, instead of aiming to measure the prevalence of all cancer survivors including those fully recovered, the aim was to measure prevalence of current cancer sufferers, i.e. those currently requiring treatment or extra care. This would give a good indication of the burden of cancer on the health services. However information on recovery times (as well as on mortality) would need to be incorporated into the model.

b) Cardiovascular disorders

Cardiovascular disorders include a range of health problems, and are classified here as follows:

Cardiovascular disease (CVD)
- Ischaemic heart disease (IHD), or coronary heart disease (CHD) (i.e. heart attack (myocardial infarction), and/or angina)
- Stroke

Other disorders
- Heart murmur, irregular heart rhythm, high blood pressure and diabetes (Colhoun, 1996)

CVD mortality typically accounts for about half of all deaths in the 35-64 age-group in industrialised countries (WHO MONICA Project, 1994). For CVD, there is a particular need for incidence data, for example, in relation to risk factors, and for studies comparing different treatment regimes.

However, for both CVD and the other cardiovascular disorders, prevalence statistics are also needed, as the numbers of sufferers are relatively large and their use of the
health services high. For example, CVD and diabetes were among the ‘top ten’ prevalent conditions included in the Dutch health status study (Ruwaaard et al, 1994). In England, a survey of adults in private households, found that about a quarter reported having had a cardiovascular disorder, the prevalence increasing with age to about half of those aged 75 or over (Colhoun, 1996).

Potential sources of data differ for the different conditions.

(i) Incidence

The incidence of CVD can be obtained from hospital records (in countries where individuals can be identified in these records), or from registers. However, national registers for CVD are less common than for cancer. One example is the recently created register for myocardial infarction in Sweden, to be followed by one for stroke. (See also section (iii) below about the WHO MONICA Project sub-national registers.)

GP records are another source for incidence data, as used in the Dutch study (Ruwaaard et al, 1994). These records have the added advantage of being sources for the incidence of cardiovascular disorders apart from acute CVD (heart attack and stroke), and for the prevalence of these disorders. They include information about medical diagnosis but will not, of course, identify cases unknown to GPs. (See also Section 6.2d.)

Surveys, as in the case of cancer, are a poor source for the incidence of acute CVD because of high mortality, non-response because of poor health and stays in hospitals and other institutions. Generally they are more suited to measuring the incidence of the chronic, rather than acute, cardiovascular conditions. The measure is, of course, subject to respondents’ knowledge of their condition and their memory of when it started.

(ii) Prevalence

GP records and surveys are currently the main sources of cardiovascular disorder prevalence data. Perhaps because of limited access to GP records, surveys are currently the most common method of estimating prevalence. The use of detailed questions, supplemented with physical measures (e.g. of blood pressure), permits estimates of both acute and chronic conditions. Surveys also permit the kinds of analysis particularly important for cardiovascular disorders, that is, analysis of certain risk factors, of health-related behaviour, and of people’s use of services and personal circumstances. A number of EU countries have carried out surveys, though methods vary greatly, so estimates are not readily comparable.

(iii) The WHO MONICA Project

The WHO ‘Multinational monitoring of trends and determinants in cardiovascular disease’ (MONICA) Project was initiated in the early 1980s. It aims to explore reasons for differences in cardiovascular mortality in different countries. It focuses on differences in the prevalence of three main risk factors, (i.e. smoking, high blood pressure, and high total serum cholesterol), and on differences in acute coronary
medical care. The project involves about 40 research centres, studying a total of about 50 populations, in 26 countries. About half of the populations are in the EU (WHO, 1989).

The project combines the use of registers and surveys. Over a ten-year period, registers are being made of people aged 35-64, suffering fatal and non-fatal coronary (and stroke) events and/or receiving acute coronary care, along with details of the coronary care received. Cross-sectional population surveys are also being carried out, at least at the beginning and end of the ten-year period, to measure prevalence of the three risk factors. While the survey methodology, and the range of information collected, has varied in each country, strict guidelines for the measurement of the risk factors are laid down.

Interestingly, the relationships between the incidence of cardiovascular disease and these risk factors, which are apparent from cohort studies, are less apparent in this study. This is thought to be because the study data on the prevalence of the risk factors comes from general cross-sectional population surveys, not specifically from the samples of disease sufferers (WHO MONICA Project, 1994). (See also Appendix 1, 1.4b.)

The project has provided a wealth of data, suitable for a wide range of analysis. However, while it reveals differences in cardiovascular morbidity between the populations studied, it covers only local/regional areas, so does not provide complete national data for comparisons.

8. **Further work on developing EU-wide comparable morbidity statistics**

Following the Task Force meeting in Luxembourg in April 1997, a meeting of experts was held in London on May 21, at which a number of suggestions were made for future work on developing EU-wide comparable disease-specific morbidity statistics. These were to:

- build on an existing inventory drawn up in 1993 to develop an inventory, in electronic form, of existing sources of data on disease-specific morbidity statistics for a limited number (say, 10) of specified diseases. As well as including information on the types of data available, the inventory should include information about meta-data and methodological issues such as the effect of mode of administration on estimates, and comments on the quality and representativeness of existing sources

- compile a list of networks and experts in Member States, and contact them to collect information on work which they have done to add value to existing sources; for example, by making disease-specific estimates from person-based data

- collect information on existing links between sources; for example, between GP records and registers, between person-based records and disease-specific records, between treatment records and person-based insurance data
• carry out collaborative methodological work using existing sources, looking, for example, at estimation methods and grossing

• develop further the approach used in Section 7 of this paper, to examine the different sources of data which are available for specific diseases

• examine some of the methodological issues involved in collecting data for ‘difficult-to-measure’ diseases such as mental illness

• encourage more Member States to join existing systems for, for example, cancer registration

• carry out pan-European pilot or feasibility studies for a common European Health Examination Survey; the results of which could be used to compare data, methodology, and calibration between Member States.
Appendix 1

Incidence and prevalence morbidity statistics, and their uses

A1.1 Incidence and prevalence measures

There are two basic types of morbidity statistic - measures of incidence and of prevalence. Incidence is a measure of the number of cases arising in a population in a period, and prevalence is a measure of the number of cases existing. For example, incidence might be expressed as the number of new cases of a disease (or disorder) per 1,000 population in a year, and prevalence as the proportion of a population with the disease at any time in a year.

Incidence may refer either to the first onset of a disease (i.e. new cases) or to all episodes.

There are two types of prevalence measure - point prevalence and period prevalence, the first representing the cases identified at a certain point in time (e.g. ‘at present’), whereas period prevalence refers to cases occurring during a specified period of time (e.g. in a particular year).

For chronic diseases/disorders (e.g. some cardiovascular conditions, diabetes, senile dementia), either ‘point’ and ‘period’ measures of prevalence are appropriate. However, for short-term diseases (e.g. influenza) or those with intermittent episodes (e.g. some mental health problems, back pain), period prevalence measures are more likely to identify all relevant cases. For prevalence statistics from different studies to be comparable, the length of period asked about must be the same.

A1.2 The relationship between prevalence and incidence

Changes in prevalence reflect changes in incidence but also in the numbers disappearing through recovery or death. The relationships between prevalence and incidence are not always straightforward. For example, increases in incidence do not always result in increases in prevalence (e.g. if recovery times are reduced). Similarly, a fall in incidence may be associated with an increase in prevalence, (e.g. if survival rates in an ageing population increase).

A1.3 Whether both prevalence and incidence measures are needed

For some diseases, particularly chronic diseases/disorders, it can be useful to have both prevalence and incidence statistics, e.g. for cardiovascular disorders, and for certain mental health problems (Ruwaaard et al, 1994; ACTA, 1989). For other diseases, one type of statistic may be the priority - for example, incidence statistics for cancer and congenital anomalies, but prevalence statistics for longer-term, less acute diseases and disorders, such as mental health, respiratory or musculo-skeletal problems.
A1.4 The uses of incidence-based statistics

Some examples of the uses of incidence-based data are described below, along with related methodological issues.

a) Monitoring the distribution of new cases (or of all episodes) among different groups in a population, including monitoring progress towards targets

Incidence data can help in understanding the distribution of diseases/disorders, and in interpreting patterns and changes in their prevalence.

In planning health care services, it can be useful to know the characteristics of new cases/episodes. Such data are most necessary for diseases requiring intensive treatment or expensive long-term care.

The relationship between the onset of one disease to the subsequent onset of other diseases can be investigated.

b) Identifying possible causes of a disease/disorder (aetiological studies)

The incidence of new cases of a disease or disorder may be related to genetic, demographic, environmental and/or socio-economic factors. To throw light on possible causes of the disease, data about clusters of new cases of the disease, about differences between groups or between regions, and about trends over time are studied.

Aetiological studies are most appropriate for diseases associated with one or more of the following: high mortality, premature death, congenital anomalies, chronic disability, long-term treatment and need for health.

Causal relationships are most likely to be identified when the data on incidence and the data on the presence or absence of possible causal factors are collected about the same sample of people, rather than from separate samples. (For example, the relationships between cardiovascular disease and various risk factors such as smoking behaviour are apparent from cohort studies which measure the incidence and the risk factors among the same people. The relationships are less apparent, however, when the incidence of disease is compared with the prevalence of the risk factors in the population as a whole, as measured in a cross-sectional survey (WHO MONICA Project, 1994)).

c) Survival analysis

Survival analysis is appropriate for diseases associated with high mortality. Survival is measured in terms of the length of time between onset of the
d) Studies of age at onset, and disease-free life expectancy

Information about age of onset, (and about the numbers of years people might expect to live prior to onset of the disease) is useful in relation to diseases associated with high mortality, (e.g. cancer) and/or chronic disability, (e.g. diabetes).

Again standard definitions are needed for the date of onset.

e) Evaluation of screening programmes

Screening programmes aimed at increasing the early detection of a disease (e.g. of breast cancer) can be evaluated with the help of incidence data. Incidence among a population screened for a disease can be compared with that among a population (of the same age) not screened. Increased incidence would generally be expected among the screened population, except where screening aids early pre-cancerous detection of abnormal cells, as is the case with cervical screening.

For diseases associated with high mortality, mortality data can be investigated to identify any consequent decrease in mortality among those screened (Quinn and Allen, 1995).

f) Analysis of alternative treatment/care regimes, and recovery rates

Information about the length of time between disease onset and recovery (or death), can be used to estimate the need for and use of health care resources by people with the disease.

Such information can also be used to investigate the effectiveness of different treatment regimes administered to different groups of patients. Strict control over the groups of patients receiving the different treatments is needed and, if comparing the results of different studies, differences between populations, definitions and methodologies should be taken into account.

A1.5 The uses of prevalence statistics

Some examples of the use of prevalence measures are as follows:
a) Monitoring the burden of disease on different groups of the population including monitoring progress towards targets

Prevalence data can help in understanding the distribution of diseases among different groups of a population at any one time. In addition, sources of data such as population surveys can collect information about several diseases, and so identify the extent of co-morbidity and measure its consequences for general health status and disability.

b) Estimating the need for and use of care services

Prevalence data can be used to estimate the need for, and use of, care from health and social services, over a period of time, by people with different diseases or disorders, and with different combinations of health problems.
Appendix 2

Examples of prevalence estimates for diseases listed in WHO HFA Target 4.6 and their validity (WHO, 1993)

A2.1 Examples of prevalence estimates derived from survey data and their validity

a) Cardiovascular disorders

Survey data appear to produce relatively accurate estimates of cardiovascular disorders. A comparison of the prevalence of self-reported cardiovascular conditions and doctors’ reports for respondents to the 1992 Health Survey for England showed a high level of agreement between the two sets of data; heart murmurs and ‘other heart trouble’ had the lowest levels of agreement. A comparison between self-reported and health examination data for the 1993 survey showed that over 90% of those who reported high blood pressure were also classified as such on the basis of blood pressure readings taken by a nurse. Among those who reported no high blood pressure, just over one in ten were assigned to the ‘high blood pressure’ group on the basis of the nurse measurements (Bennett et al, 1995).

b) Cancer

Survey data are not so useful for acute, high fatality or short-duration illnesses such as some types of cancer. Estimates are biased by high mortality, the social stigma which can arise from the disease and non-response because of poor health and hospital stays. Prevalence is low (usually estimated at 1-1.5%), which means that very few cases will be identified in an individual survey, making only limited analysis possible. People with cancer may also be unaware of the true nature of their condition (Egidi, 1996). The Health Survey for England has identified congenital anomalies and ‘neoplasms’ as particular areas of difficulty for the coding of diseases to ICD classifications (White et al, 1993).

c) Psychiatric conditions

Psychiatric conditions tend to be under-reported in open questions asking about long-standing conditions (General Household Survey 1972). Surveys specifically designed to measure psychiatric morbidity offer the opportunity of comparing self-reported data with data on medication and with clinical examinations (Meltzer et al, 1995a)

d) Diabetes

A comparison of estimates of self-reported doctor-diagnosed diabetes with glycated haemoglobin (GHB) levels by the Health Survey for England showed that only a small proportion (2-3%) of non-diabetic respondents had a raised GHB - 8% or more). (Colhoun et al, 1996).

e) Respiratory conditions
Self-reported chronic respiratory conditions can also be compared with the results of tests of respiratory function (1995 HSE).

f) **Allergies**

Data on allergies from the 1994 Danish Health and Morbidity Survey showed that about two thirds of those reporting allergies had had their allergy diagnosed; less than half had undergone clinical tests.

**A2.2 Examples of prevalence estimates derived from GP data**

The England and Wales GP Research Database has produced estimates of the prevalence of diseases targeted by the *Health of the Nation* report (Department of Health, 1992), some of which are also the subject of the WHO HFA Target 4.6. They are coronary heart disease, hypertension, schizophrenia, depression and anxiety, and asthma (all of these treated in the last year), stroke in the last year in patients still alive, schizophrenia ever diagnosed, and diabetes both insulin-dependent and non-insulin dependent) (ONS, 1996b).

The Netherlands has produced prevalence estimates of the ‘top ten’ diseases, both from their continuous database and from the GP survey. Interestingly, with the exception of depression, there is no overlap with the ‘top ten’ diseases in terms of incidence (Ruwaard et al, 1994).
References


Danish Institute for Clinical Epidemiology. Questionnaire for the Danish Health and Morbidity Survey 1994 with response frequencies. DICE, Copenhagen.


This report was produced by a contractor for Health & Consumer Protection Directorate General and represents the views of the contractor or author. These views have not been adopted or in any way approved by the Commission and do not necessarily represent the view of the Commission or the Directorate General for Health and Consumer Protection. The European Commission does not guarantee the accuracy of the data included in this study, nor does it accept responsibility for any use made thereof.