Chapter 6. Items of patient information which may be collected by registries

R. MacLennan

Queensland Institute of Medical Research, Bramston Terrace, Herston, Brisbane, QLD 4006, Australia

The information needed by a cancer registry is directly related to and determined by its functions. Hospital registries are primarily concerned with surveillance of cancer patients from a hospital, and they are discussed separately in Chapter 13. Only population-based registries can accurately assess the incidence of cancer in the general population. Although the items of information required can be completely specified only after the functions and purpose of a registry have been considered (see Chapters 3 and 4), there is a set of basic items common to almost all registries.

Basic items of information for cancer registries

Many items of information which are essential for a registry concerned with patient management are clearly not essential for a population-based registry primarily concerned with the estimation of cancer incidence. The term 'basic information' is used for those items that are generally collected by all cancer registries. Whether or not other items are collected will depend on the purpose of the registry, the method of information collection (see Chapter 5), and on the resources available to the registry. It is important to distinguish between items collected by a registry and items stored by a registry—not all items collected are stored in coded form (e.g., items of information used for administrative purposes).

The basic items of information for any cancer registry are listed in Table 1. Many cancer registries have foundered because they have attempted to collect too much information. The emphasis must be put on the quality of the information collected rather than quantity. Some of the most successful and productive registries only collect a very limited amount of information for each patient. These items of basic information are relevant for population-based registries everywhere; they might be the only items to be collected by registries in developing countries (see Chapter 14).

Optional items of information

Each additional item of information increases the complexity and cost of registration. Thus, for each additional item the registry should ask 'Why do we need it?' and 'Can we afford the cost of collecting it?', rather than 'Would we like to have it?'. A comprehensive list of items is given in Table 2.

Table 1. Basic information for cancer registries

Item no.	Item	Comments
The person		
	Personal identification[a]	
3	Name	According to local usage
4	Sex	
5	Date of birth or age	Estimate if not known
	Demographic	
6	Address	Usual residence
11	Ethnic group[b]	When population consists of two or more groups
The tumour		
16	Incidence date	
17	Most valid basis of diagnosis	
20	Topography (site)	Primary tumour
21	Morphology (histology)	
22	Behaviour	
35	Source of information	E.g., hospital record no., name of physician

[a] The minimum collected is that which ensures that if the same individuals are reported again to the registry, they will be recognized as being the same person. This could also be a unique personal identification number

[b] Ethnic group is included here because it is important for most registries, especially in developing countries

For specialized research registries such as digestive tract, or childhood tumour registries, the basic items may be added to in a modular fashion, with collection and coding of the additional modular information being the responsibility of the specialist users.

Population-based registries do not often undertake active follow-up of patients, since they are not concerned with assessing response to therapy. However, they can assess the overall survival rate of patients with different forms of cancer, which is the least ambiguous measure of outcome. In order to do so, they must collect information on date of death of registered cases (see Chapter 12).

Items collected on samples of patients

Obviously, a cancer registry cannot collect everything on everybody. Core information may be supplemented with *ad hoc* information from samples, so that studies can be undertaken which otherwise would not be feasible. Samples may be defined by person, tumour type, or time. Thus the collection of non-melanoma skin cancers may be limited to, say, three-year periods every ten years. Special *ad hoc* studies of certain tumours may require collection of an extended range of data items for limited periods of time. Measures of quality of life of a series of cancer patients may be feasible only on a sample.

National and international comparability of items

At the national level, the definitions of items and the codes used by a cancer registry should accord with those used in other systems. Thus, for instance, demographic

Table 2. Items of information which may be collected

Item no.[a]		Item
The person		
Identification		
1	(2)	Index number
2	(3)	Personal identification number
3	(4)	Names
Demographic and cultural items		
4	(5)	Sex
5	(6)	Date of birth
6	(8)	Address
7	(7)	Place of birth
8	(9)	Marital status
9	(11)	Age at incidence date
10	(52)	Nationality
11	(54)	Ethnic group
12	(53)	Religion
13	(55, 56)	Occupation and industry
14	(77)	Year of immigration
15	(78)	Country of birth of father and/or mother
The tumour and its investigations		
16	(13)	Incidence date
17	(17)	Most valid basis of diagnosis of cancer
18	(81)	Certainty of diagnosis
19	(57)	Method of first detection
20	(18)	Site of primary: topography (ICD-O)
21	(19)	Histological type: morphology (ICD-O)
22	—	Behaviour
23	(21)	Clinical extent of disease before treatment
24	(23)	Surgical-cum-pathological extent of disease before treatment
25	(59)	TNM system
26	(60)	Site(s) of distant metastases
27	(20)	Multiple primaries
28	(64)	Laterality
Treatment		
29	(22, 65–70)	Initial treatment
Outcome		
30	—	Date of last contact
31	(24)	Status at last contact
32	(25)	Date of death
33	(26, 76, 84)	Cause of death
34	(83)	Place of death
Sources of information		
35.1	—	Type of source: whether death certificate, physician, laboratory, hospital or other
35.2	—	Actual source: name of laboratory, hospital, physician, etc.
35.3	—	Dates

[a] Item numbers in parentheses refer to the equivalent item(s) in the *WHO Handbook for Standardized Cancer Registries* (WHO, 1976a)

codes (population groups, occupation, residence etc.) should be identical with those of the census and statistics bureaux that supply denominators for epidemiological analysis.

Because of the need to have national comparability between the numerators collected by a registry and available denominators, full international standardization is not feasible. For example, race and country of origin of immigrants may be defined differently in, say, the United States of America and in Australia. For such items of information, international comparability can be achieved by the methods of collection of data and calculation of rates, and not necessarily in the detailed nomenclature of individual items. The details and extent of international comparability will thus vary. They can and must be greatest for the description and coding of tumours (see Chapter 7). Recommendations in this chapter may be considered as a basis for comparability for those variables unlikely to form part of the census data that will be used for population denominators. They are based on current practices of cancer registration throughout the world.

Fixed and updatable items

Fixed items are those which cannot be modified in the light of subsequent information, for example, the clinical extent of disease before treatment (item 23). This does not apply to errors, which must be corrected. *Updatable items* are those which can be modified in the light of new information, e.g., the most valid basis of diagnosis of cancer (item 17). This needs to be distinguished from information collected as part of cancer patient follow-up, where both the old and the new information may be included in the patient's data together with dates, e.g. multiple primaries (item 27).

Personal identification

Unambiguous personal identification (items 2 and 3) is essential in all cancer registries. It is needed to prevent duplicate registrations of the same patient or tumour and to facilitate various functions of cancer registries, such as obtaining follow-up data and performing record linkage. It is more important that, for a given region, sufficient identifying information be available than that the actual specific items be internationally standardized. The emphasis should be on adequate personal identification rather than on the specific items that contribute to personal identification, since these vary considerably from one country to another. An identification number or social security number exists in many countries, and may be very useful for patient identification.

Other personal characteristics are described which are useful for personal identification and as independent descriptive parameters in relation to cancer, e.g., date of birth and sex. Precise date of birth is one of the most valuable items of personal identification and it should always be recorded if available. Approximate age is sufficient to describe cancer patterns.

Description of the neoplasm

This central aspect of cancer registration includes anatomical site (item 20), morphology (item 21), behaviour (item 22), multiple primaries (item 27), pretreat-

ment extent of disease (items 23–25) and most valid basis of diagnosis of cancer (item 17). Anatomical site is the most common axis for tabulations. Its coding in a special adaptation of the International Classification of Diseases for Oncology (ICD-O) differs from the coding of topography in the current edition of the International Classification of Diseases (ICD-9). Classification and coding of neoplasms are described in detail in Chapter 7.

Pretreatment extent of disease is described by items 23–25, which relate to two aspects of the extent of disease in the initial phases of diagnosis and therapy. The first, commonly referred to as clinical staging (item 23), pertains to the extent of disease, as assessed clinically, before the initiation of any treatment. The TNM classification (item 25) (Hermanek & Sobin, 1987) is often used. The second, referred to as surgical-pathological staging (item 24) contains information on extent of disease that is available from initial surgical therapy, and includes histological information about lymph node involvement etc., or information from autopsy, if the patient died before treatment could be given.

Description and coding of items of patient information

Items of patient information are described systematically below, with a definition of each item and comments on its relevance. Each item may have several categories or classes. Although coding is an input operation (see Chapter 8), it is more convenient to give suggested codes here with the description of the items. The coding of neoplasms is complex and is discussed separately in Chapter 7. Coding is complicated by changes in classifications. Payne (1973) describes the practical problem: 'Committees responsible for the design of national and international classifications and codes cause some inconvenience to cancer registries and similar organizations by too frequent changes. When such changes take place registries may either follow them but only from the time changes become effective, or they may convert the coding of all existing records to conform to the changes. In the former case awkward discontinuities persist in the registry's data which complicate analyses extending over a long period; in the latter case, the conversion process may be time-consuming, expensive and possibly liable to introduce systematic errors.'

A set of numbers has been assigned to the recommended data items. The previous numbering system which was proposed for hospital tumour registries (WHO, 1976a) and was included in the publication *Cancer Registration and its Techniques* (MacLennan *et al.*, 1978) is shown in parentheses in Table 2.

The person

IDENTIFICATION

Item 1: Index number

A registration number is assigned by the cancer registry to each patient. This number is given to all documents and items of information relating to the patient. If a patient has more than one primary tumour (item 27), each tumour is given the same registration number. These primary tumours can be distinguished by site (item 20),

morphology (item 21) and incidence date (item 16). This question is discussed further under item 27. Use of a patient registration number rather than a tumour registration number is recommended, as this facilitates the analysis of multiple primaries and simplifies patient follow-up. One widely used numbering system includes the last two digits of the year in which the patient first registered, together with a serial number for the year. For example, |8|7|0|0|0|0|1| is the registration number given to the first patient registered in 1987. The second patient registered in 1987 would be given the number |8|7|0|0|0|0|2|.

The year of registration may be different from the year in which the patient was first admitted to hospital and diagnosed. For instance, a patient admitted and diagnosed in October 1986 may not be registered until January 1987. In this case, the registration number will begin with 87, although the year for calculation of incidence will be 1986, as reflected in the incidence date (item 16).

This is discussed in further detail in Chapter 8.

Item 2: Personal identification number

Many countries use a personal identification number that is unique to an individual; it may incorporate other personal information, such as date of birth and sex. Some countries have no such personal identification number; others have more than one. Examples include the national identity number in Nordic countries, Malaysia and Singapore, and the social security number in the USA.

The utilization of these identification numbers in medical records varies greatly. They are more likely to be available when they serve an administrative purpose associated with medical treament or hospital admission or with providing benefits to patients. If a suitable number is available for only a very few patients, then it should not be relied on for patient identification, but whenever such a number exists, the cancer registry should promote its inclusion in the hospital files (preferably at the time of admission). In countries where identification numbers are ubiquitous, they can also serve as the index number (item 1).

The complete number should be obtained, including any check digits when these exist. It must be noted that the number as written may be incorrect—transposition of digits occurs commonly, or another person's number may be written on a form.

Item 3: Names

The full name is essential for identification in cancer registries. Although this item appears to be simple to obtain, there may be many problems with names, especially in developing countries. It is recommended that names be copied from identity cards whenever possible.

Spelling of names. There are often different spellings for names with the same pronunciation, e.g., Reid and Read, Petersen and Pedersen. With regard to unwritten languages and dialects in developing coutries, subtle distinctions in sound may not be expressed by the phonetic system used for medical records (English, French, Spanish), and the same name may be spelled differently on different occasions. Ambiguities owing to spelling can be greatly reduced by use of a special code system, e.g. the New York State Identification Intelligence System (NYSIIS) (see Appendix 3c).

Abbreviations. Persons often, but not consistently, use abbreviations of names, e.g., the name James may be modified to Jim or Jimmy, Robert to Rob, Bob, Bobby etc.

Titles. Titles can be used to assist identification, although they may not be used consistently, e.g., Doctor, Father, Mother, Brother, Sister in certain religious orders, and Mrs, Miss or Ms in some English-speaking countries.

Changes of names. Name changes during a person's lifetime may considerably complicate the registry's task. A common example is in societies in which women change their family name following marriage. In many non-industrialized societies, names are changed at other stages of life. In many developing countries, additional information may be available, including affiliation, i.e., the father's name; in many Latin American countries the mother's family name is often given on documents. Most registries will need to make provision for the recording of multiple names—particularly recording of maiden name for married women who take their husband's name.

Order of names. Conventions vary as to the order in which names are written. In western European cultures, the family name may be written either first or last, depending on the context, although in everyday speech the family name is stated last. In many parts of Asia, the family name is invariably given first. The order in which names are written should be standardized for each registry and should reflect local practice.

DEMOGRAPHIC AND CULTURAL ITEMS

Item 4: Sex

Sex is a further identifying item and is invariably found in hospital records; however, in many other sources of information, the sex may not be recorded. Although sex may be inferred in some cultures from the given name or from the wording of the hospital summary, in others it is not easy to determine, for example, in reports of cancer that are based on pathology reports only. Persons who change their phenotypic sex by means of operations and drugs should be coded separately. Suggested codes are:

1. Male	3. Other
2. Female	9. Unknown

Item 5: Date of birth

Date of birth is of great importance in assisting identification, particularly when there is limited variation in names, or when other specific identifying information is lacking. The related item, age at date of tumour incidence (item 9), may be derived from the date of birth (if known). Alternatively, if the date of birth is not known, the year of birth may be estimated in years from the approximate age. This is useful in constructing birth cohorts. The date of birth on an identity card may be the result of a guess, but, provided it is used consistently on all documents, it is useful for identification.

For international comparability, it is necessary to convert any local dating system

or convention to the standard system used internationally, for instance, by United Nations organizations. The date should be recorded in clearly labelled boxes:

| 25 | May | 1933 |

Day Month Year

The international convention is to write these in the order illustrated. (The reverse order has certain advantages in data processing, but this can readily be achieved electronically.) Day and year should be given in figures in full, and month in words; this will avoid ambiguities such as occur in data from the USA.

There is a distinction between the recording of a date and its coding: thus, the date written above is coded as:

| 2 | 5 | 0 | 5 | 1 | 9 | 3 | 3 |

Item 6: Address

Address is useful in patient identification. It is also essential to identify the registered cases who are residents of the registry area, in order to calculate incidence rates and to study variations in incidence by subregion of the registry area. Address is also required if any follow-up of cases is to be carried out.

The address recorded should be the patient's usual residence, and this must be distinguished from his or her address at the time of entering hospital. If there is an identity card, this will normally give the patient's usual residential address, which may be copied by the hospital. This could help to distinguish residents from another area who are staying temporarily with relatives. Patients may intentionally give an address in the area served by a specific hospital in order to qualify for acceptance or free treatment by that hospital. In some areas, identity cards may be borrowed for the same purpose.

For population-based registries, the place of usual residence should be coded, using the same classification and codes as those used for available population denominator data. The most detailed codes available should be used, in order to minimize the effects of changes in administrative or political boundaries within a country or region. Thus detailed codes can be regrouped to conform to new boundaries or definitions of denominator data (e.g., urban or rural, or postal codes in the UK and USA). Identification of non-resident patients is important and they must be excluded from incidence (and survival) studies. Unless this is done there may be considerable distortion, particularly if the registration area contains a treatment centre of renown. Thus, the Tata Memorial Hospital treats head and neck cancers from all over India; inclusion of all patients would exaggerate the importance of cancers at these sites in the Bombay population.

Item 7: Place of birth

Place of birth may assist in personal identification, and it may provide clues to cancer etiology. Studies of persons who move from one environment to another may show differences in cancer incidence in the two environments. Such movement can be studied between countries, e.g., Japan and the USA, and also within countries.

Whenever possible, the precise birthplace in the country of origin should be recorded, since national boundaries may have changed; e.g., an individual born in 1910 in Breslau, then in Germany, would now live in Wroclaw, Poland. The year of immigration (item 14) is also of interest in studies of migrants (see below).

Codes used should conform to those used in national vital statistics. These may include places within the geographical area covered by the cancer registry in addition to places in other parts of the country or in other countries. An example of the latter is the list of geocodes for place of birth used by the US SEER program. WHO also has a list of codes of countries for international use (see Appendix 1). The same considerations apply to nationality (item 10), ethnic group (item 11) and religion (item 12).

Item 8: Marital status

Although it may be used as an item of personal identification, it must be remembered that marital status may change during the course of an illness. This item is widely available and the registry is fairly certain to get accurate information in a high proportion of cases. The codes adopted should preferably be those used in vital statistics.

Item 9: Age at incidence date

This refers to the age in years at the incidence date (item 16). In many populations, age may not be known accurately, or may deliberately be stated inaccurately. If the date of birth is unknown, the year of birth may be estimated from the stated age and recorded as a fixed item (item 5). In many cases, it is not essential to record age but it acts as a useful check upon birth-date.

Age is of great relevance in the description of cancer incidence, but precise age is not essential. In developing countries, approximate age may be estimated in a number of ways (Higginson & Oettle, 1960); for example, the person may have married at the time of some event whose date is known. The Chinese and other groups in Asia follow a system in which the names of animals are assigned to different calendar years of birth in a 12-year recurring cycle. Thus, among Chinese, it may be possible to validate reported age if the animal year of birth is also recorded.

Item 10: Nationality

For most purposes, nationality is equivalent to citizenship, which is defined as the legal nationality of a person. There may be difficulties in obtaining accurate information about stateless persons, persons with dual nationality and other ambiguous groups. Nationality must be distinguished from place of residence (item 6) and place of birth (item 7).

If census information on nationality is available, rates may be calculated, in which case the definitions of nationality as used by the census bureau should be used by the registry.

Item 11: Ethnic group

This is considered to be an essential item for many cancer registries. Social and cultural differences between groups may be related to the utilization of medical

facilities and to the acceptance of programmes for early detection. Ethnic group may be an indicator of differences in culture and habits which determine exposure to carcinogenic factors, since different ethnic groups may differ in occupational specialization, diet and other habits and customs. Information on subgroups within major ethnic groups may also be important, particularly for providing clues to etiology. Thus, in Singapore, information on the occurrence of cancer in various distinct groups speaking Chinese dialects has revealed important differences in cancer patterns (Lee *et al.*, 1988).

The ethnic characteristics about which information is needed in different countries depend on national circumstances. Some of the bases on which ethnic groups are identified are: country or area of origin, race, colour, linguistic affiliation, religion, customs of dress or eating, tribal membership or various combinations of these characteristics. In addition, some of the terms used, such as 'race' or 'origin', have a number of different connotations. The definitions and criteria applied by each registry for the ethnic characteristics of cancer cases must, therefore, be determined in relation to the groups that it wishes to identify. By the nature of the subject, these groups will vary widely from country to country, so that no internationally standardized criteria can be recommended.

Because of the interpretative difficulties that may occur, it is important that when this item is recorded the basic criteria used be defined. The definitions of ethnic groups used by cancer registries should be compatible with official definitions used for census reports, but may need to be more detailed. Even if no population census figures are available, information on ethnic group is important for proportionate morbidity analyses.

A problem may arise when an ethnic group is disguised for political or other reasons; this is the case with the Chinese in certain South-East Asian countries. They are not distinguishable on the basis of routine medical records, and documentation of their cancer patterns would need a special survey.

Items 12: Religion

The optional collection of information on religion as a separate item will depend on local conditions: the number of religions, feasibility of collection and possible relevance. Religion may determine the attitude towards, and the use of, modern medical services and thus influence knowledge about malignant disease. Women in some religious groups are reluctant to use medical services (especially those involving examination by male physicians), and their true cancer incidence may thus be grossly under-estimated.

Religious beliefs may directly affect exposure to carcinogens or may be an indicator of cultural differences which affect exposure. The cancer patterns in religious groups in the USA, particularly Mormons and Seventh Day Adventists, have been a fertile source of hypotheses relating cancer risk to dietary and other lifestyle factors.

Information on religion may be incorporated into the definition of ethnic group (item 11).

Item 13: Occupation and industry

Occupation refers to the kind of work done by an employed person (or performed previously, in the case of unemployed or retired persons), irrespective of the industry or of the status of the person (as employer, employee etc.). An example might be: a lorry driver in transport or mining industries, or in government.

Industry refers to the activity of the establishment in which an economically active person works (or worked) (United Nations, 1968). Some occupations are specific to an individual industry. The International Labour Office (ILO, 1969) has published the *International Standard Classification of Occupations, 1968* (ISCO), and a new edition is currently under preparation. The United Nations (United Nations, 1968) has published the *International Standard Industrial Classification of all Economic Activities* (ISIC). These classifications were created primarily for economic purposes and are thus often inadequate for studies of cancer. The *Classification of Occupations and Directory of Occupational Titles* (CODOT), published by the Department of Employment, UK (DOE, 1972) gives more specific details of occupation, which are more relevant to potential exposure; it will soon be replaced by a new Standard Occupational Classification, which will be compatible also with the ISCO (Thomas & Elias, 1989).

Information on occupation is frequently poorly reported to registries. Often, the status at the time cancer occurred is reported, which may be irrelevant to the occupational status some 20 to 30 years previously; the latter is more significant in relation to possible etiology. Nevertheless, although it must be treated with caution, even imperfect information of this kind may be of value.

Population-based cancer registries play an important role in studies of occupational cancer risk by providing an economical follow-up mechanism for cohort studies (see Chapter 3).

Item 14: Year of immigration

This is of interest for registries dealing with migrant groups including immigrant workers. Relating the date of incidence to the date of immigration permits the study of the effects of duration of residence in the new environment on the risk of cancer, or alternatively, the effect of age at the time of migration on the change in risk. In a country with many migrants, e.g., Israel, a sudden rise in the incidence rates of cervical cancer soon after an immigration period could be ascribed to an increase in diagnosis rate rather than to a real increase in the incidence of this disease in the population group.

Item 15: Country of birth of father and/or mother

The country of birth of the parents may be of interest in countries with sizeable immigrant populations. In the USA, the study of changes in risk in first and second generations of Japanese migrants has been of particular interest in evaluating the importance of environmental changes (particularly in relation to diet). In these cases ethnic group and place of birth serve to distinguish first and second generation migrants. In other countries, only place of birth of parents may allow identification of

second generation migrants. Usually, country of birth of father is more often available in denominator data.

The tumour and its investigations

Item 16: Incidence date

This is not necessarily the date of first diagnosis by a physician, as this may be difficult to define precisely. For patients seen in hospital, it is the date of first consultation at or admission to a hospital for the cancer—and this includes consultation at outpatient departments only. This is a definite point in time which can be verified from records and is the most consistent and reliable date available throughout the world. For these reasons, it is chosen as both the anniversary date for follow-up and survival computation purposes and as the date of onset for measuring incidence, henceforth referred to as the incidence date.

If the above information is not available, other dates may have to be used. Thus, incidence date refers to, in decreasing order of priority:

(*a*) date of first consultation at, or admission to, a hospital, clinic or institution for the cancer in question;

(*b*) date of first diagnosis of the cancer by a physician or the date of the first pathology report—a population-based registry should seek this information only when necessary for recording the incidence date;

(*c*) date of death (year only), when the cancer is first ascertained from the death certificate and follow-back attempts have been unsuccessful; or

(*d*) date of death preceding an autopsy, when this is the time at which cancer is first found and was unsuspected clinically (without even a vague statement, such as 'tumour suspected', 'malignancy suspected').

If there is a delay between first consultation and admission for definitive treatment, the date of first consultation at the hospital is selected (both consultation and treatment may be outpatient; for example, in nasopharyngeal carcinoma). If cancer is diagnosed during treatment for another illness, e.g., a person being treated for a chronic disease develops symptoms during inpatient or outpatient treatment and cancer is detected, the appropriate incidence date is the date of diagnosis.

A special problem is posed by cases known to the registry only from death certificates. If the registry does not succeed in obtaining further information but nevertheless includes such cases (see item 19), the general rule is to take the date of death as the date of incidence.

Item 17: Most valid basis of diagnosis of cancer

The information of greatest interest for the assessment of reliability of incidence rates is the most valid method of diagnosis used during the course of the illness. The most valid basis of diagnosis may be the initial histological examination of the primary site, or it may be the post-mortem examination (sometimes corrected even at this point when histological results become available). This item must be revised if later information allows its upgrading.

When considering the most valid basis of diagnosis, the minimum requirement of

a cancer registry is differentiation between neoplasms that are verified microscopically and those that are not. To exclude the latter group, as some pathologists and clinicians might be inclined to do, means losing valuable information; the making of a morphological (histological) diagnosis is dependent upon a variety of factors, such as age, accessibility of the tumour, availability of medical services, and, last but not least, upon the beliefs of the patient and his or her attitude towards modern medicine.

A biopsy of the primary tumour should be distinguished from a biopsy of a metastasis, e.g., at laparotomy; a biopsy of cancer of the head of the pancreas versus a biopsy of a metastasis in the mesentery. Cytological and histological diagnoses should be distinguished.

Morphological confirmation of the clinical diagnosis of malignancy depends on the successful removal of a piece of tissue which is cancerous. Especially when using endoscopic procedures (bronchoscopy, gastroscopy, laparoscopy, etc.), the clinician may miss the tumour with the biopsy forceps, despite seeing it. These cases must be registered on the basis of endoscopic diagnosis and not excluded through lack of a morphological diagnosis.

Care must be taken in the interpretation and subsequent coding of autopsy findings, which may vary as follows:

(*a*) the post-mortem report includes the post-mortem histological diagnosis;

(*b*) the autopsy is macroscopic only, histological investigations having been carried out only during life;

(*c*) the autopsy findings are not supported by any histological diagnosis.

For coding, methods of diagnosis have been divided into two broad categories, non-microscopic and microscopic, each consisting of four further categories. These are given below in approximate order of increasing validity. With advances in diagnostics, expansion of these codes to two digits may be considered, keeping the overall principle of distinguishing non-microscopic and microscopic diagnoses.

Non-microscopic

1. Clinical only
2. Clinical investigation (including X-ray, ultrasound etc.)
3. Exploratory surgery/autopsy
4. Specific biochemical and/or immunological tests

Microscopic

5. Cytology or haematology
6. Histology of metastasis
7. Histology of primary
8. Autopsy with concurrent or previous histology
9. *Unknown*
10. *Death certificate only*—if no other appropriate code is available, as in registries which use only the basic data set (see Table 1). Registration on the basis of information included in the death certificate alone (item 35.1), for which no other information can be traced, must be distinguished from cases first coming to the registry's attention by means of a death certificate mentioning cancer and where diagnosis is based on other information.

Item 18: Certainty of diagnosis
It may be useful to include a code to express the certainty of the coded diagnosis in addition to the most valid basis. Even the pathologist, in making an autopsy report, may be unable to state the origin of the tumour but may give a choice of two or three possibilities. This item could be used to indicate doubts as to the stated histological diagnosis or, on the other hand, to express confirmation after revision by a specialist. To some extent, uncertainty as to diagnosis is expressed by the use of a topography code (item 20) 199.9 (unknown primary site) and morphology codes (item 21) 8000 (neoplasm, tumour, malignancy, cancer) and 9990 (no microscopic confirmation of tumour), and by behaviour code (item 22) 1 (uncertain whether benign or malignant).

Other codes which may be used separately from the ICD-O coding system are:
1. Malignancy uncertain, site uncertain
2. Malignancy uncertain, site certain
3. Malignancy certain, site uncertain
4. Malignancy certain, site certain, histology uncertain
5. Histological diagnosis doubtful after revision
6. Histological diagnosis confirmed after revision
7. Malignancy certain, site certain, histology certain

Item 19: Method of first detection
The evaluation of data from time series will often be easier if information exists on the method or circumstance whereby cancer cases are first diagnosed in the population. This information is different from the most valid basis of diagnosis (item 17) and it refers to the means by which the cases came to medical attention. In particular, the introduction of screening programmes may influence incidence rates by the diagnosis of prevalent cases in the preclinical phase, some of which would never have progressed to symptomatic cancer. Cancers first detected by autopsy examinations can also be identified, so that the extent to which incidence rates—for example of cancer of the prostate—reflect extensive autopsy-detected cases can be evaluated.

Suggested codes are:
1. Screening examination
2. Incidental finding (on examination, at surgery)
3. Clinical presentation (with symptoms)
4. Incidental finding at autopsy
8. Other
9. Unknown

Item 20: Site of primary tumour: topography (ICD-O)
The detailed topography of a tumour is the most important item of data recorded, and it provides the main axis of tabulation of registry data.

In abstracts from clinical records, the location of the tumour should be written in words, with as much specific information as possible, i.e., with the full clinical diagnosis; for instance, 'primary malignant neoplasm of left upper lobe of lung', 'malignant tumour of colon, hepatic flexure', 'metastatic tumour in lung, primary unknown'. The information for this item should be updated whenever additional data

become available, e.g., in the last example, the primary site may subsequently be reported, leading to a change in the coding of topography (but not in the incidence date, item 16).

As described in detail in Chapter 7, registries are strongly recommended to use the special International Classification of Diseases for Oncology (ICD-O) (WHO, 1976b). In this case, the topography code which should be used refers to the anatomical location of the primary tumour.

With ICD-O, topography is coded regardless of the behaviour of the tumour. Benign tumours and tumours of undefined behaviour are thus given the same topographical code as malignant neoplasms. Thus, a code for behaviour (benign, *in situ*, malignant) must be used in addition. This may be the fifth digit of the morphology code (see item 21 below), but if morphology is not coded, a special behaviour code (item 22) is necessary.

Item 21: Histological type: morphology (ICD-O)

Although the anatomical site of a tumour is the usual axis for the reporting of cancer registry data, the importance of detailed morphology is being increasingly recognized, and not only as an index of confidence in the diagnosis. In the past, lymphomas, leukaemias, melanomas and choriocarcinomas were the only malignant morphological diagnoses that could be identified in the International Classification of Diseases (ICD). However, morphology is often related to etiology and prognosis and, hence, must be considered in many epidemiological and clinical studies. An unusual histological type may be the first indication of a new environmental carcinogen, e.g., angiosarcoma of the liver following exposure to vinyl chloride. The choice of therapy and assessment of prognosis are influenced by the histological type.

The complete histological diagnosis, as stated in the pathology report, must be recorded by cancer registries. The registry may decide to record the laboratory reference number (see item 35.2) which may facilitate future access to the blocks used to make histological sections or to the slides themselves for review purposes.

The wording of the histological diagnosis may pose problems in coding. Even for a common tumour, the diagnosis of which would give rise to no dispute, terminology may differ according to various schools. It would be of great help if pathologists could be persuaded to use the terms of the ICD-O morphology chapter. For a detailed discussion of the coding of morphology see Chapter 7.

The ICD-O should be used universally for describing morphology, even by registries that continue to code anatomical site by the standard ICD. Indeed, the index of the ninth revision of the ICD contains the ICD-O morphology codes (WHO, 1977).

Item 22: Behaviour

If morphology is coded using ICD-O, the fifth digit expresses behaviour of the tumour (see Chapter 7). For registries which do not include the histology code in their database (and all are strongly urged to do so), behaviour would be recorded separately using the following ICD-O conventions:

0. Benign
1. Uncertain whether benign or malignant
   Borderline malignancy

2. Carcinoma *in situ*
   Intraepithelial
   Non-infiltrating
   Non-invasive
3. Malignant

Notice that, in keeping with the recommendations made in Chapter 7, behaviour codes 6 and 9 of the ICD-O should **not** be used by cancer registries.

Item 23: Clinical extent of disease before treatment
Item 24: Surgical-cum-pathological extent of disease before treatment

The staging of cancer has a long tradition. Staging of different cancers is important in planning treatment, indicating likely prognosis, evaluating the results of therapy, and facilitating exchange of information between treatment centres. These functions are mainly related to clinical practice, hence careful recording of the extent of disease is an important role of the hospital-based cancer registry. Population-based registries will, in general, be less able to record accurate or consistent information on the extent of disease for all cases registered. Stage of disease in a population-based registry may be used to provide information on the timing of diagnosis (as an indication of public awareness of the significance of signs and symptoms of cancer, or the result of programmes of early detection), or as a means of ensuring comparability in studies of population-level survival (Hanai & Fujimoto, 1985).

Extent of disease may be recorded as both the clinical extent (reflecting the clinical opinion of the doctor at the time of diagnosis) and surgical-cum-pathological extent, in which clinical observation is augmented by the findings at surgery (including microscopic examination), if this is part of the initial treatment, or the findings at autopsy, if the patient died before treatment could be given. *In practice, if they decide to record it at all, population-based registries will normally have only one item for 'extent of disease' based on the maximum amount of information available at the time of treatment*. Extent of disease recorded during follow-up is not of interest to population-based registries.

A variety of staging schemes have been proposed for solid tumours. Some are specific to certain cancer sites, such as the FIGO staging system for gynaecological cancers (American College of Obstetricians & Gynecologists, 1973), and the Duke's system for colo-rectal cancers; others are applicable to all tumour types. The most detailed of these latter schemes is the TNM system, described below (Item 25), but there are several more compact schemes. An example, taken from the Summary Staging Guide of the SEER program (Shambaugh *et al.*, 1977) is given below, together with suggested codes:

0. *In situ*
1. Localized
2. Regional: direct extension to adjacent organs or tissues
3. Regional: lymph nodes
4. Regional: direct extension *and* regional nodes
5. Regional: NOS

6. Distant (non-adjacent organs, distant lymph nodes, metastases)
7. Non-localized, NOS
8. Not applicable
9. Unknown (or not staged)

Use of this scheme requires that, for each site, lymph nodes which are considered 'regional' or 'distant' be defined; such definitions are provided in the publication cited (Shambaugh *et al.*, 1977), in the American Joint Committee on Cancer's *Manual for Staging of Cancer* (Beahrs *et al.*, 1988), and in the TNM system (see Item 25). For reporting of results, some grouping together of the above categories is required, and many registries record extent of disease as a simpler 'summary staging' scheme, such as the following example:

—*In situ*
—Localized
—Regional
—Distant

The American Joint Committee on Cancer provides for a similar summary grouping; in this, localized tumours are subdivided into two groups (I and II) on the basis of their size, information which may prove difficult for a population-based registry to obtain.

None of the generalized staging systems described above is appropriate for recording the extent of lymphomas. These are generally categorized into four stages; for tumours of lymph nodes and lymphoid tissue, the stages and their definitions are:

1. Localized: *one* lymphatic region above or below the diaphragm
2. Regional: more than one lymphatic region on *one side* of the diaphragm
3. Distant$_1$: lymphatic regions on *both sides* of the diaphragm
4. Distant$_2$: disseminated involvement of one or more extralymphatic organs

Item 25: TNM system

The TNM classification of cancers at various sites is now well established on an international basis (Hermanek & Sobin, 1987).

The TNM system has three main components. 'T' represents the extent of the primary tumour, with suffixes to differentiate the size of the tumour or involvement by direct extension. 'N' indicates the condition of the regional lymph nodes, with suffixes to describe the absence or increasing degrees of involvement by tumour. 'M' indicates the presence or absence of distant metastases. Additional features of each field can be indicated by subscripts, e.g., microscopic findings.

TNM provides a very detailed categorization, which for most purposes is readily condensed into the summary stages (*In situ*, Local, Regional, Distant) described above. For further discussion, the reader is referred to Davies (1977).

Population-based registries which receive data from multiple sources must be aware of difficulties of comparing TNM staging from hospital to hospital.

Item 26: Site(s) of distant metastases

Although this is a low-priority item for population-based registries, clinicians frequently ask for it to be included. As described in Chapter 7, the ICD-O topography

code allocated as item 20 should refer only to the site of the primary tumour in cancer registration. If it is wished to collect information on the site of metastases, space could be allocated for several ICD-O topography codes; however, this degree of detail is rarely required and a simple one-digit code is preferable. Suggested codes are:

0. None
1. Distant lymph nodes
2. Bone
3. Liver
4. Lung/pleura
5. Brain
6. Ovary
7. Skin
8. Other
9. Unknown

Item 27: Multiple primaries

There are many problems with the term 'multiple primaries'. More than one tumour may occur at different sites in the same organ or in different organs, with the same or different histology and at the same or different times. The registry's medical coder must decide if multiple tumours are manifestations of a single neoplasm, i.e., one primary with metastasis, or if they are different primary tumours. The registry must have clear procedures for the classification and coding of multiple primary tumours. Definitions used for the registration and reporting of multiple primary cancers are given in detail in Chapter 7.

Multiple primary tumours may be identified by means of a suffix (2, 3 etc.) to the index number (item 1), as proposed in Chapter 8. This avoids the need for a special field to indicate second and subsequent tumours, a solution which requires cross reference to the index number of the first tumour. Alternatively, a separate tumour number can be used in addition to a personal identification number (item 2) (see Chapter 8).

Item 28: Laterality

In paired organs, such as lung, the side involved may be important in the choice of therapy. In other cases (e.g., retinoblastoma, nephroblastoma) unilateral and bilateral tumours have different etiological significance. The paired organs for which laterality codes are to be used must be defined by the registry. Appropriate codes are:

1. Right
2. Left
3. Bilateral
9. Unknown

Treatment

Item 29: Initial treatment

For the population-based registry, this item should be initial treatment, started within four months of first diagnosis. Since treatment practices vary from place to place, and even within one centre in the course of time, it is advisable to collect data in very broad categories.

Provision should be made for the identification of patients who did not receive

initial tumour-directed treatment, since such persons are important for survival studies and for studies of the natural history of the disease.

Population-based registries should aim to collect as little information as possible in this category—perhaps just a summary of the objectives of therapy, e.g.:

1. Symptomatic only
2. Palliative only
3. Curative—incomplete
4. Curative—complete
5. Uncertain
7. Other
8. No treatment
9. Unknown

Often, however, clinicians concerned with the work of the registry will insist that the nature of therapy, and the date on which therapy commenced, are specified. In this case, a grouping of codes for nature of the initial therapy might be:

0. No treatment (or symptomatic only)
1. Surgery
2. Radiotherapy
3. Chemotherapy
4. Immunotherapy
5. Hormonotherapy
8. Other therapy
9. Unknown

A decision as to how to code procedures such as cryotherapy, laser treatment etc., should be reached.

Several treatment modalities may have been used, and the registry may decide to code all those used in a defined period (e.g. four weeks after first treatment), together with dates of starting.

Outcome

Item 30: Date of last contact

The date at which the patient was last known to be alive may be known from follow-up visits to hospital, by contacting the patient's medical attendant, or from the patient. This date is important if survival rates are to be computed (see Chapter 12).

At the time of registration, date of last contact should be set equal to incidence date (item 16), unless additional information such as hospital discharge date etc. is available. It is then updated when further contacts become known to the registry. If the patient dies, date of last contact could be deleted or, preferably, made identical to date at death (item 32).

Item 31: Status at last contact

Population-based registries may only be able to obtain information as to whether the patient is alive or dead. To go further requires active follow-up of patients, an activity more characteristic of hospital registries. This item is also essential for computation of survival.

Suggested codes are:

1. Alive
2. Dead
8. Emigrated
9. Unknown

Hospital registries will wish to elaborate on category 1. Alive, specifying, for example, whether there was evidence of tumour presence or not.

Item 32: Date of death
The complete date of death, including day, month and year, should be recorded to facilitate tracing of death certificates and other information relating to the individual. This item enables computation of survival.

Item 33: Cause of death (ICD)
Two options are available. As a minimum the registry may use the codes:

1. Dead of this cancer
2. Dead of other cause
9. Unknown

This enables the corrected survival rate to be calculated, as described in Chapter 12. Alternatively, the registry may record the ICD code appropriate to the actual cause of death, if this has been determined by personnel experienced in determining underlying cause from death certificates. The coding of cause of death can be very complex since this embraces the full range of the ICD and involves the application of specific rules for the allocation of underlying cause. Special training is therefore needed. If registry staff are required to code cause of death, they should be trained in national vital statistics offices, and periodic checks must be made on the validity of their coding. The population-based registry will often know only that death has occurred and have no information on the cause, e.g., non-medical certification of death. If death certificates are received from national vital statistics offices, they may already be coded according to the ICD.

It should be noted that this item is not used to determine which cases are first notified to the registry by means of a death certificate (item 19), or those registered on the basis of death certificate information only (items 17 or 35.1).

Item 34: Place of death
This information may be useful for both hospital and population-based registries. No codes are proposed, but should be developed by a registry to reflect local practice, e.g., death at home, in a hospice, in hospital, etc.

The population-based registry may use this information as an indication of certain aspects of medical care, e.g., a tendency to discharge terminal patients in order to diminish the number of deaths in hospital statistics.

Sources of information

A registry needs a comprehensive coding scheme incorporating all sources of information used by the registry (see Chapter 5). Thus for cases notified from a hospital it would include hospital code, date of admission or discharge, and hospital number. For cases notified from a laboratory, the scheme would have laboratory code, date of biopsy (or its receipt), biopsy number or laboratory reference number.

The hospital record number can facilitate reference back to hospital files for

additional information not included in the cancer registry. When separate records are kept, the hospital department may also have to be identified and coded.

As the same patient may be reported by several hospitals, a population-based registry or a hospital registry serving several hospitals will have to code each hospital in addition to the record number.

'Death certificate only' (DCO) cases are defined as those for which no other information concerning the patient can be traced even after approaching the hospital or clinician responsible for completing the death certificate. These cases may be dealt with separately when the registry's data are analysed. As described in Chapter 9, the proportion of such DCO cases provides a useful indication of quality control for registries. If possible a record should also be kept of those cases which first come to the registry's attention from a death certificate mentioning cancer. The percentage of such cases is a useful guide to the adequacy of case-finding mechanisms (Chapter 5). These cases should, however, not be confused with DCO cases defined above (see also item 17).

These items can also facilitate the administrative aspects of a registry by documenting the source of the data. The following categories are proposed as a provisional guide, but individual registries should devise their own scheme.

Item 35.1. Type of source
Hospital, laboratory, primary care physician, death certificate alone, or other.

Item 35.2. Actual source
Name of laboratory, hospital or doctor; laboratory reference number, etc.

Item 35.3. Dates