Chapter 7. Classification and coding of neoplasms

C. S. Muir[1] and C. Percy[2]

[1]*International Agency for Research on Cancer, 150 cours Albert-Thomas, 69372 Lyon Cédex 08, France*
[2]*National Cancer Institute, National Institutes of Health, Bethesda, MD 20892-4200, USA*

Introduction

Classification of neoplasms involves their arrangement or distribution in classes according to a method or system. Neoplasms can be classified in many ways but, for clinician and cancer registry alike, the two most important items of information are the location of the tumour in the body (synonyms: anatomical location; site; topography) and the morphology, i.e., the appearance of the tumour when examined under the microscope (synonyms: histology, cytology), as this indicates its behaviour (malignant, benign, *in situ*, and uncertain). Cancer registries endeavour, as a minimum, to classify each neoplasm according to its topography, morphology and behaviour, as well as recording particulars of the host.

Sound classification requires an agreed nomenclature—a series of names or designations forming a set or system—so that, for example, all histopathologists agree to give a particular microscopic appearance the same name.

The custodians of a classification have a three-fold task: first, to ensure that the classification adapts to accommodate changes in concepts and user needs, otherwise the classification will fall into disuse; second, to ensure that such changes as are made avoid the inclusion of terms and concepts that are ephemeral, and third, to ensure that changes are made in such a way as to permit continuity of time series.

For convenience, most classifications assign numerical codes to their constituent entities so that a frequently complex series of pieces of information can be conveyed, stored and retrieved in the form of numbers. With the continual advances in electronic and computer techniques, it is possible today to eliminate manual coding and enter the descriptors directly, letting the computer assign code numbers, but this added convenience does not influence the basic concepts of disease classification.

At first glance the classification and coding systems currently used seem illogical and needlessly complex. This is due, in part, to the fact that cancer is but one of many diseases and is thus assigned a niche in the larger classification systems which have developed over time. The principal manual for classifying diseases is the International Classification of Diseases (ICD) published by the World Health Organization, the ninth revision of which (WHO, 1977) is in current use. It will be

described in detail below, but first it is useful to have some knowledge of the evolution of the classifications used, as this helps to explain their current format and structure.

Historical review of topographical and morphological classifications of neoplasms (1948–1985)

An excellent history of disease classification prior to 1948 is given in the introduction of ICD-7 (WHO, 1957). After the United Nations was established following the second world war, WHO was created as a specialized United Nations agency dealing with health, and took over the responsibility for the International Lists of Causes of Death. In 1948, WHO published the sixth revision of ICD (WHO, 1948) and the classification has been revised usually every 10 years thereafter (see Figure 1).

Chapter II of the ICD, dealing with neoplasms, is primarily a topographic classification arranged according to the anatomical site of the tumour, except for a few histological types such as melanomas, lymphomas and leukaemias. Basically the structure of the neoplasms chapter has not changed for the past 40 years. Neoplasms were allotted 100 consecutive three-digit code numbers running from 140 to 239. These numbers are also commonly called categories or rubrics. From ICD-6 onwards most organs (or categories) have also been subdivided with a fourth digit giving greater anatomical detail, e.g., in ICD-7, 141.0 was assigned to malignant neoplasms of the base of the tongue. Organs were arranged according to organ systems, for example ICD-7 rubrics 150-159 covered the malignant neoplasms of digestive organs and peritoneum. Neoplasms with a given behaviour were grouped into blocks designated malignant, benign, and of unspecified nature; beginning with ICD-9, blocks were also allotted to *in situ* neoplasms and to neoplasms of uncertain behaviour. The structure of ICD-9 is illustrated by the example in Table 1.

In the 1940s, the first cancer registries had already recognized the need for distinguishing between histologically different tumours of the same organ (Clemmesen, 1965). A histological classification of tumours was not furnished in ICD-6, which, for example, provided no way to distinguish between a squamous cell

Table 1. Structure of chapter II, neoplasms, of the International Classification of Diseases, Ninth Revision (ICD-9) categories 140–239

Behaviour of neoplasms	Organ systems	Organ site	Organ subsites
Malignant (140–208)	Buccal cavity, pharynx (140–149)		
	Digestive system (150–159)	Oesophagus (150.-)	
		Stomach (151.-)	
		Small intest. (152.-)	
		Colon (153.-)	Hepatic flexure (153.0)
		Etc.	Transverse colon (153.1)
			Descending colon (153.2)
			Etc.

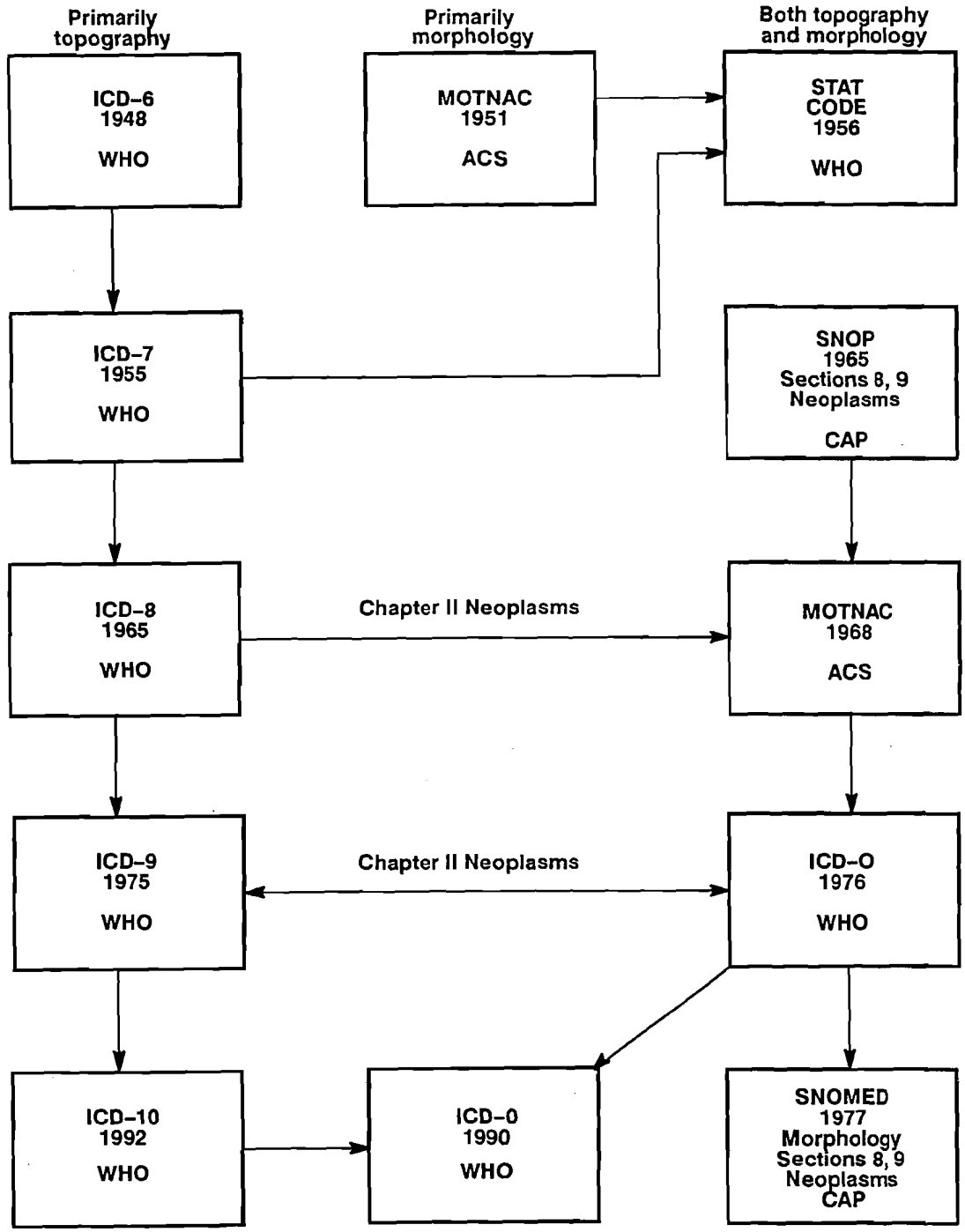

Figure 1. Codes for neoplasms 1948–1985
WHO, World Health Organization; ACS, American Cancer Society; CAP, College of American Pathologists; ICD, International Classification of Diseases; MOTNAC, Manual of Tumor Nomenclature and Coding; STAT, Statistical Code for Human Tumours; SNOP, Systematized Nomenclature of Pathology; SNOMED, Systematized Nomenclature of Medicine.

carcinoma of the lung and an adenocarcinoma of the lung; both were classified as malignant neoplasm of lung (ICD-6 162) (and still are in ICD-9). Therefore, in 1951, the American Cancer Society (1951) developed and published its first *Manual of Tumor Nomenclature and Coding* (MOTNAC). This had a three-digit morphology code, of which the first two digits gave histological type and the third the behaviour of

the tumour. Cancer registries at that time usually used the malignant neoplasm section of ICD-6 for coding topography and MOTNAC for morphology. This principle was later adopted by WHO when in 1956 it published a *Statistical Code for Human Tumours* (WHO, 1956), which consisted of a topography code based on the malignant neoplasms chapter of ICD-7 (WHO, 1957) and the morphology, including behaviour code, of MOTNAC (see Figure 1).

The College of American Pathologists (1965) published the Systematized Nomenclature of Pathology (SNOP). This included a two-digit (and a highly detailed four-digit) topography code to cover all anatomy (not just cancer sites) and a morphology code, of which sections 8 and 9 were assigned to neoplasms. In addition there were four-digit codes for the fields of etiology and function. It was agreed that the American Cancer Society could use sections 8 and 9 from SNOP for the morphology section of a revised MOTNAC, which appeared in 1968 (Percy *et al.*, 1968). The revised MOTNAC had no relation to the original 1951 edition. Instead the topography section was based on the topographic structure of the malignant neoplasm section of ICD-8 (WHO, 1967) (see Figure 1), while the four-digit morphology code provided (behaviour being the fourth digit) was taken from SNOP.

When the ninth revision of ICD was being developed, WHO asked the International Agency for Research on Cancer (IARC) to make recommendations concerning the content and structure of the neoplasms chapter (Chaper 2) in consultation with the Cancer and ICD units of WHO in Geneva. In the course of this work, the worldwide need for a logical, coherent and detailed classification for neoplasms was recognized. Thus, a working party was formed that developed the International Classification of Diseases for Oncology (ICD-O) (WHO, 1976b), which categorized a tumour by the three axes of topography, morphology and behaviour. The topography section was based on the malignant neoplasms chapter of ICD-9, the morphology field on MOTNAC (Percy *et al.*, 1968), which was expanded by one digit (from three to four), and finally a behaviour code following a slash or solidus (/). In addition, a grading code (degree of differentiation) was provided as the sixth digit of morphology.

At the same time, the College of American Pathologists (1977) revised SNOP as the Systematized Nomenclature of Medicine (SNOMED). SNOMED incorporated the ICD-O morphology section for its morphology sections 8 and 9—Neoplasms. The SNOMED topography section on the other hand, as in SNOP, has no relation to ICD-9 or ICD-O topography, since it covers all anatomical structures and not just the sites where tumours occur.

Classification and coding

A cancer registry is faced with a number of problems when deciding on the classification to be used for the coding of tumours. These include the degree of detail desirable, internal comparability of long time series (a particular problem for existing registries) and international comparability between registries.

The underlying principle of coding is to bring together in classes cases of cancer which have common characteristics. While classification by etiology, prognosis and response to treatment would be highly desirable, such information is frequently

obtained some time after diagnosis. Based on current knowledge tumours are still best delineated on the three axes of site of tumour, histopathological appearance and behaviour. The cancer registry should therefore code its tumours by an internationally accepted system, using all three axes, which easily allows the classification of tumours in more or less broad categories.

ICD-9 fulfils many of the requirements, but lacks the logic, flexibility and histological detail of ICD-O, which is recommended for use in cancer registration. SNOMED shares many of the advantages of ICD-O, but lacks the international recognition attached to the ICD classification system. Although revision of SNOMED is planned by its publisher, the College of American Pathologists, only ICD and ICD-O will therefore be described in detail in the following pages.

International classification of diseases, 1975 revision (ICD-9) (WHO, 1977)

The ICD-9 manual is published as two volumes: Volume 1 gives a numerical listing; Volume 2 an alphabetical index. The manual is designed for the coding and classification of both mortality (death certificates) and morbidity (hospital and other medical diagnoses). A United Nations treaty engages 44 nations to code and report mortality from their countries using the current ICD, but the treaty does not include cancer registry data. Several rules for the coding of morbidity are included in the back of Volume 1, in addition to those dealing with the choice of underlying cause of death.

In ICD-9, the neoplasms chapter comprises the categories (rubrics) running from 140 to 239 inclusive. These rubrics are further divided as follows into six groups according to the behaviour of the neoplasms.

Categories	*Group*
1. 140–199	Malignant neoplasms (other than those of lymphatic and haematopoietic tissue)
2. 200–208	Malignant neoplasms of lymphatic and haematopoietic tissue
3. 210–229	Benign neoplasms
4. 230–234	Carcinoma *in situ*
5. 235–238	Neoplasms of uncertain behaviour
6. 239	Neoplasms of unspecified nature

The greatest anatomical detail is provided for the malignant neoplasms. Most three-digit rubrics are further subdivided by means of a fourth digit.

Although in essence topographical in axis, ICD-9 includes several morphological categories, sometimes mixed with topography, e.g., the distinction of malignant melanoma of skin (ICD-9 172) from the other forms of skin cancer (ICD-9 173). For several rubrics the axis is a tissue, no matter where located, e.g., connective and soft

tissue, or lymphatic and haematopoietic tissue. The complete list of malignant neoplasms of such "morphological" rubrics is as follows:

Malignant neoplasm of connective and other soft tissue: ICD-9 171
Melanoma of skin: ICD-9 172
Malignant neoplasm of placenta (choriocarcinoma): ICD-9 181
Hodgkin's disease: ICD-9 201
Non-Hodgkin lymphoma: ICD-9 200, 202
Multiple myeloma: ICD-9 203
Leukaemias: ICD-9 204-208

While the benign neoplasms (ICD-9 210–229) are also classified for the most part on grounds of anatomical location, several of the rubrics are morphological or relate to a connective or other soft tissue:

Lipoma: ICD-9 214
Other benign neoplasm of connective and other soft tissue: ICD-9 215
Uterine leiomyoma: ICD-9 218
Haemangioma and lymphangioma, any site: ICD-9 228

The diagnosis of carcinoma *in situ* (ICD-9 230–234) can only be made microscopically, as the critical feature is the lack of invasion of the malignant cells through the basement membrane of the epithelial tissue involved. Such neoplasms are classified topographically.

The neoplasms of uncertain behaviour (ICD-9 235–238) are those with a well defined histological appearance, but whose subsequent behaviour is difficult to forecast, e.g., granulosa cell tumours of the ovary (ICD-9 236.2).

The index of ICD-9 also contains all the morphological (histological) codes of the morphology (M) field of ICD-O (see below).

International Classification of Diseases for Oncology (ICD-O), first edition (WHO, 1976b)

ICD-O is an extension or supplement of the neoplasms chapter, i.e., Chapter II, of ICD-9. It permits the coding of all neoplasms by:

(*a*) topography (T) (four digits),
(*b*) histology (morphology) (M) (five digits) including behaviour (one digit following a /) i.e., malignant, benign, *in situ*, uncertain whether malignant or benign; and
(*c*) one digit for grading (grades I–IV) or differentiation (well differentiated to anaplastic).

A tumour is thus completely characterized by a ten-digit code, e.g., a well differentiated adenocarcinoma of the lung is coded as T-162.9 M-8140/31 (lung 162.9, adenocarcinoma 8140, malignant behaviour /3, well differentiated 1).

Topography

All topographic categories have the same code number within the range 140 (Lip) through 199 (Unknown site) as ICD-9 except for the categories 155.2: Liver, not specified whether primary or secondary, 172: Malignant melanoma of skin, and 197.-[1]: Secondary malignant neoplasms of respiratory and digestive systems, and 198.-: Secondary malignant neoplasms of other specified sites. These categories were not used since they could be handled in ICD-O by using the behaviour codes /6 (metastases), or /9 (uncertain whether primary or metastatic site), or by using the site category 173 for skin in conjunction with the morphology code numbers 8720/3-8780/3 which denote one of the forms of malignant melanoma. (It will be recalled that, in ICD-9, rubric 173 denotes 'Other malignant neoplasms of skin', i.e., those that are not malignant melanomas).

ICD-O contains a code number, 169, which does not appear in ICD-9. This provides a topographic point of reference for malignant neoplasms of the reticuloendothelial and haematopoietic systems, i.e., those neoplasms which would be coded to ICD-9 rubrics 200–208.

ICD-O	169.-	Haematopoietic and reticuloendothelial system
	169.0	Blood
	.1	Bone marrow
	.2	Spleen
	.3	Reticuloendothelial system
	.9	Haematopoietic system

Since histogenetically the spleen fits here, the ICD-9 code for spleen, 159.1, was dropped from ICD-O.

The meaning of the ICD-9 rubric 196, Secondary and unspecified malignant neoplasm of lymph nodes, was changed in ICD-O topography to permit the coding of primary tumours of the lymph nodes, this number being used in ICD-O as the topographic site for both Hodgkin's and non-Hodgkin lymphomas. A lymphoma originating in an organ would be coded to the relevant T-category. Thus, a malignant lymphoma of the stomach would be coded in ICD-O as T-151.9 M-9590/39 and a gastroenterologist could include it in a series of stomach tumours. Using ICD-9, such tumours would be coded 202.8, i.e., the same code as for a nodal lymphoma and their organ of origin would be lost. Since 20–25% of all non-Hodgkin lymphomas are extranodal and considered different from those arising in lymph nodes, the ability to code such neoplasms separately is an important feature of the ICD-O system.

Morphology

In order to encompass different classifications accepted by pathologists, the authors of ICD-O and its predecessor MOTNAC decided to assign code numbers to

[1] When there is more than one fourth digit within the rubric and it is not wished to or it is not possible to code any particular one, the convention is to use the first three digits followed by a dot (.) and a dash (-). The former recognizes the existence of fourth digits, the latter that no specific one has been coded.

all terms appearing in the major classification schemes for tumours. For example, Hodgkin's disease can be classified according to both the largely obsolete Jackson–Parker classification (Jackson & Parker, 1944) (M-9660/3 to M-9662/3) and the Lukes–Collins (Lukes & Collins, 1974) or Rye classification (Lukes & Butler, 1966) (M-9650/3 to 9657/3). The inclusion of six international classification schemes for non-Hodgkin lymphoma in the original ICD-O makes its use complicated for these tumours, but gives it a large degree of flexibility. With the advent of the working formulation in 1982 (National Cancer Institute, 1982; Percy *et al.*, 1984) and the updating of the lymphoma section of the ICD in the second edition of the ICD-O (Percy *et al.*, 1990), the coding of the current classifications has been clarified.

Some examples illustrating the above points are given below in Table 2.

The WHO series *International Histological Classification of Tumours* (WHO, 1967–1978) was used as a basis for selecting preferred terms in ICD-O. This series—the so-called Blue Books—was initially developed by international committees between 1967 and 1978. These monographs represent the opinions of leading specialists throughout the world and now comprise a series of 26 volumes, one for each major site or system of neoplasms. The books are profusely illustrated and colour slides may be purchased. Initially there was no coding scheme, but with the advent of ICD-O, the relevant morphology code numbers were added in Volume 22. In 1978, WHO prepared a summary of these histological entities: a compendium of the first 20 books (1967–78) of this series (Sobin *et al.*, 1978). This gives the histological terms used for each site (for Blue Books Nos. 1–26) with the corresponding ICD-O code number. Several of these classifications have now been revised.

Behaviour

This is the fifth digit of the morphology code and is used to distinguish between benign and malignant neoplasms and the stages in between: *in situ* and uncertain whether malignant or benign, as well as primary and metastatic sites.

The codes are:

/0 Benign
/1 Uncertain whether benign or malignant
  Borderline malignancy
  Low malignant potential
/2 Carcinoma *in situ*
  Intraepithelial
  Non-infiltrating
  Non-invasive
/3 Malignant, primary site
/6 Malignant, metastatic site
  Secondary site
/9 Malignant, uncertain whether primary or metastatic site

Table 2. Coding of selected cancers according to ICD and ICD-O

Term	ICD-9	ICD-O (First edition)		ICD-10	ICD-O (Second edition)[b]	
Malignant melanoma of skin	172	T-173.-	M-8720/3 to M-8780/3	C43.-	C44.-	M-8720/3 to M-8790/0
Hodgkin's disease	201	T-196.-[a]	M-9650/3 to M-9662/3	C81.-	C77.-[a]	M-9650/3 to M-9667/3
Non-Hodgkin lymphoma	200,202	T-196.-[a]	M-9590/3 to M-9642/3 M-9690/3 to M-9722/3 M-9740/3 to M-9750/3	C82-C85.- " " " "	C77.-[a]	M-9590/3 to M-9595/3 M-9670/3 to M-9714/3
Multiple myeloma	203	T-169.-	M-9730/3 to M-9731/3	C90.-	C42.1	M-9731/3 to M-9732/3
Leukaemia	204–208	T-169.-	M-9800/3 to M-9940/3	C91-C95.-	C42.1	M-9800/3 to M-9940/3

[a] If not extranodal
[b] See section on ICD-10 and ICD-O (Second edition) below

Grading or differentiation

This, the sixth and final digit of the morphology code, has five categories which are:

1	Grade I	(Well) differentiated
2	Grade II	Moderately (well) differentiated
3	Grade III	Poorly differentiated
4	Grade IV	Undifferentiated, anaplastic
9	Grade or differentiation not determined, not stated or not applicable	

The appropriate differentiation codes are included with each grade, for example, Grade I and well differentiated. This code is useful, since a clinician's decision about management of a patient may hinge on information about whether a tumour is stated to be well differentiated or anaplastic. Thus, for instance, gynaecologists may decide on different treatments for well differentiated endometrial carcinoma (panhysterectomy with or without post-surgical irradiation) and for anaplastic endometrial carcinoma (presurgical irradiation). However, "the use of grading varies greatly among pathologists throughout the world, and in many instances malignant tumours are not routinely graded" (WHO, 1976b).

Use of ICD-O

The structure and use of ICD-O are carefully outlined in the introduction to ICD-O and will not be repeated here. It is important that cancer registries using the ICD-O familiarize themselves with the conventions.

An explanation of a few items that are of importance in the application of ICD-O to the cancer registry setting are outlined below, as well as items which experience has shown provide particular difficulties.

Matrix system

The ICD-O matrix is explained in the introductory pages of that classification (page xix). Nevertheless, this tends to create problems when programming in computerized registries. Potentially, nearly any epithelial tumour can have an '*in situ*' phase, but only about six morphological types with *in situ* are listed specifically in ICD-O. The behaviour code /2 (i.e., *in situ*) can be attached to any of the four-digit morphology code numbers for solid tumours if the *in situ* form exists and is diagnosed, e.g., papillary adenocarcinoma *in situ* is coded 8260/2. Provision must be made in the computer programs for these terms so that they are not flagged as errors. This type of problem may also arise for a tumour that usually is benign, but is stated by the pathologist to be malignant. While it is useful to have a flag to draw attention to such an occurrence, once the diagnostic statement is verified the tumour must be accepted and included. (The reverse may also occur, i.e., a tumour which is usually malignant but has been diagnosed as benign).

No microscopic proof

It is not advisable to attribute a morphology to a tumour which has not been microscopically examined. The morphology code M-9990 in ICD-O was provided for

users wishing to denote that a tumour had not been microscopically confirmed. Almost all registries will code in addition whether the diagnosis had a microscopic basis, was a clinical diagnosis, based on X-ray, etc. Such a field is usually called basis of diagnosis (see Item 17, Chapter 6).

Primary site and the behaviour code in ICD-O

The amalgamation of information on behaviour (malignant, *in situ*, unknown) and on origin (primary site, metastatic site, unknown) for a given tumour in one behaviour code poses a potential problem for the use of ICD-O by cancer registries. Tumour registries should primarily identify tumours by the topographic site where the tumour originated—in other words, the primary site—and tabulations should be made by primary site. To help identify the primary site in ICD-O, the behaviour code /3 means malignant, primary site. If for some reason the primary site is unknown, but the disease is certainly malignant, the code T-199.9 M—/3 should be used (T-199.9 is the code for unknown primary site.) Sometimes it is clear that there are metastases to, for example, the lungs or liver, but the true site of origin of the tumour cannot be determined. This case should also be coded to T-199.9 M—/3 unknown primary site.

Although tumour registries prefer not to have a large number of cases assigned to unknown site, it is better to know that the specific categories are "clean".

The ICD-O makes provision for site-specific morphology terms. Some morphological types of neoplasm are specific to certain sites, e.g., nephroblastoma (8960/3) to kidney, and basal-cell carcinoma (8090/3) to skin. For these morphological types, the appropriate topography number has been added in parentheses. It is suggested that, for these morphological types, the site-specific topography term can be coded if a site is not given in the diagnosis. However, if a site is specified, then this should be coded, even if it is not the topography proposed. For example, the site-specific T-number, T-174.- (female breast) is added to the morphological term Infiltrating duct carcinoma, because this term is usually used for a type of carcinoma which arises in the breast. However, if the term Infiltrating duct carcinoma is used for a primary carcinoma arising in the pancreas, the correct T-number would be 157.9 (pancreas, NOS).

Coding of metastases

ICD-O provides for coding the presence of a metastasis in a given organ with a behaviour code /6, but this facility should *not* be used in tumour registries (behaviour code /9—uncertain whether primary or metastatic site—is therefore also redundant). The topography code will refer only to primary site (see above).

The /6 code for behaviour was designed for use by pathologists who receive, for example, tissue from the lung or liver, look under the microscope and recognize a metastasis but do not know where the tumour originated. A pathology laboratory would code this T-162.9 (lung) and M—/6 meaning metastasis from some other organ to lung. Although a tumour registry could follow the same convention, by not doing so, it solves the coding problem posed when the primary site is known but the tumour is histologically diagnosed on the basis of a metastasis. For example, a surgeon may choose to remove a lymph gland close to the stomach rather than taking a biopsy from

the primary gastric cancer. In such circumstances, the cancer registry should code the primary site, namely stomach, including the morphology of the metastasis, with behaviour /3. If the registry wishes to distinguish between tumours verified by microscopic examination of the primary cancer and those confirmed from histological examination of a metastasis, an additional code specifying the basis of the diagnosis should be used (see Chapter 6, item 17). If, for example, a tumour is reported as being clinically a primary carcinoma of the lung and the diagnosis is supported by microscopic examination of mediastinal lymph nodes showing metastatic squamous-cell cancer, it should be coded as T-162.9 (lung), M-8070/3 (squamous-cell carcinoma). The basis of diagnosis code would in this instance be 6, i.e., histology of metastasis.

Using this convention, the information on the site of the metastasis from which a biopsy was taken is lost. However, registries wishing to collect information about the sites of distant metastases are better advised to do so using a separate variable Site(s) of distant metastases (see Chapter 6, item 26).

Advantages and disadvantages of ICD-9 and ICD-O

In this discussion, the various points made concerning the relative merits of ICD-9 and ICD-O are for the most part applicable to ICD-10 and the second edition of ICD-O (see below).

ICD-9

The major advantage of the ICD is that it is truly international, being used by all WHO Member States for tabulation of causes of death and for most health statistics. This is an advantage which outweighs all drawbacks. However, for the cancer registry, the combination of axes of classification within a single code number does raise problems, e.g., ICD-9 rubric 172, malignant melanoma of skin, conveys information on three axes: malignancy, organ affected, and histological type. However, other malignant tumours of skin are assigned to ICD-9 rubric 173 where, although the fourth digit allows for coding of various parts of the body surface, it is not possible to code the clinically more important distinction between basal-cell and squamous-cell carcinomas. Indeed, for the majority of sites, no separation of histological types is possible in ICD-9. It will be recalled that the index for ICD-9 contains all the morphological terms of the ICD-O, and hence it would be quite feasible for cancer registries to assign the usual ICD-9 code number and add the ICD-O morphology code. To do so loses much of the advantage to be derived from adding histology. Hodgkin's disease of the stomach would be coded 201 (Hodgkin's disease) followed by M 9650/3 (Hodgkin's disease). The use of ICD-O would result in T-151.-, M-9650/3, thus preserving the location of the lesion. For cancer registries, it is essential that histology is coded. ICD-O should therefore be used. It is a relatively simple task to convert ICD-O to ICD-9 if so needed. Although some specialities have complained that for certain anatomical sites the topographic subdivisions provided in ICD-9, and hence ICD-O, are not sufficient, it is suggested that extra digits should be confined to special studies. The Dental Adaptation of ICD-8 (WHO, 1978) is a good

example of a well constructed topographic expansion, collapsible into the parent ICD.

ICD-O

The major advantage of ICD-O is its logic and detail which provide optimal facilities for coding and reporting. The degree of detail is often believed to render its use difficult. On the contrary, experience shows that the degree of detail and the index of synonyms make it easy to locate the correct code number and minimize the judgements often involved in the use of less detailed coding schemes. The detailed coding of each tumour provides an excellent basis for the construction of conversion tables to less detailed codes. Also, childhood cancers should for the most part be classified according to histology rather than topography, and an international classification scheme for childhood cancer has been based on the morphology and topography codes of the ICD-O (Birch & Marsden, 1987).

Retrieval and tabulation of data coded by ICD-O are more complex than for ICD-9 or ICD-10. For registries storing their data in a computer-readable form, this should not prove a major difficulty.

ICD-O, like ICD, is truly international, having been made available in eight languages: English, French, German, Italian, Japanese, Portuguese, Russian and Spanish. It has gained widespread acceptance, being used in both hospital and population-based registries. Some 76 registries contributing to Volume V of the series *Cancer Incidence in Five Continents* use ICD-O (Muir *et al.*, 1987).

Implementation of use of ICD-O by cancer registries

New registries

Any cancer registry beginning operations can implement use of ICD-O and should record both topography and morphology (including behaviour and grading of tumours), using the second edition of ICD-O (Percy *et al.*, 1990).

Established registries

Registries that have used ICD or any other coding scheme with or without a histology classification (e.g., MOTNAC) may consider changing to ICD-O. As mentioned above, the degree of detail in ICD-O makes it possible to maintain continuity with regard to topography for long time series. Computerized cancer registries may consider coding by ICD-O, incorporating a conversion table in the registration program for automated coding to the current revision of the ICD. Further information on conversions is given in the section on tables of ICD conversions below.

ICD-10 and ICD-O second edition (Percy et al., 1990)

As noted earlier, the ICD is revised every 10 years or so. The 10th Revision will come into operation on 1 January 1993. Given the need for ever-greater detail and for the recognition of new diseases and syndromes, it was decided that the number of three-

digit categories available in ICD-9 was insufficient to permit useful expansion. The 10th Revision of ICD will thus be alphanumeric, not numeric, and will provide about 2000 categories at three-digit level, of which neoplasms have been allotted 150. Malignant neoplasms are assigned to C00 to C97, *in-situ* neoplasms D00–D09, benign neoplasms D10–D36 and neoplasms of uncertain and unknown behaviour D37–D48.

The order of existing fourth digits has occasionally been changed. Thus for colon, some fourth digits in ICD-9 have been given three-digit status in ICD-10, e.g., rectosigmoid junction (C14), and several new entries have been created, notably for mesothelioma (C45), Kaposi's sarcoma (C46), malignant neoplasm of peripheral nerves and autonomic nervous system (C47), and malignant neoplasm of soft tissue of retroperitoneum and peritoneum (C48). The section on non-Hodgkin lymphoma has been completely revised (C82–C85), a rubric created for malignant immunoproliferative disease (C88) and for multiple independent primary neoplasms (C97). ICD-10 also provides a series of rubrics for the coding of human immunodeficiency virus (HIV) disease. One of these (B21), displayed below, is of particular interest to cancer registries:

B21 *Human immunodeficiency virus [HIV] disease resulting in malignant neoplasms*

- B21.0 HIV disease resulting in Kaposi's sarcoma
- B21.1 HIV disease resulting in Burkitt's lymphoma
- B21.2 HIV disease resulting in other non-Hodgkin lymphoma
- B21.3 HIV disease resulting in malignant neoplasms of lymphoid, haematopoietic and related tissue
- B21.7 HIV disease resulting in multiple malignant neoplasms
- B21.8 HIV disease resulting in other malignant neoplasms
- B21.9 HIV disease resulting in unspecified malignant neoplasm

The ICD-10 coding rules for determination of underlying cause of death are such that several malignant neoplasms will be assigned to rubric B21, i.e., outside the neoplasms chapter, in mortality statistics, and cancer registries undertaking death clearance or searching hospital discharge diagnoses will need to examine records for deaths or admissions ascribed to this rubric. It will be obvious from the content of the rubric B21 that unless the registry has access to the certificate or case records, the anatomical location or nature of some neoplasms coded to B21 will be 'lost'.

In parallel with the development of the neoplasms chapter of ICD-10, the opportunity was taken to update ICD-O, notably in the area of malignant neoplasms of lymphatic, haematopoietic and related tissues (see Table 2). A small number of obsolete terms have been discarded and new terms and synonyms added. Hydatidiform mole, NOS is considered a benign neoplasm, as in the first edition, and neurofibromatosis including Von Recklinghausen's disease, except of bone, to be a neoplasm of unknown and uncertain behaviour. These terms in ICD-10 are coded to **O**01.9 and **O**85 respectively. The second edition of ICD-O was published in 1990 (Percy *et al.*, 1990). Although the 10th Revision of ICD does not enter into force until 1 January 1993, WHO has given permission for the second edition of ICD-O to use

the rubrics C00-C97 for topography in conjunction with the revised morphology codes and cancer registries may wish to consider its use as from, say, 1 January 1991.

Multiple tumours

It has long been recognized that a given individual may have more than one cancer in his or her lifetime. With increasing survival after treatment for several forms of cancer, and the use of chemotherapeutic agents which are themselves carcinogenic in the treatment of malignant disease (Schmähl & Kaldor, 1986; Day & Boice, 1983), it is estimated that at present some 5% of all cancer patients develop a further independent primary cancer (Flannery *et al.*, 1983; Storm & Jensen, 1983).

As most registries count tumours, not patients, it is highly desirable to have a series of rules to define the circumstances under which an individual is considered to have more than one cancer. Although every tumour registry has the prerogative to set its own rules, it should pay attention to the comparability of its data with those of other registries as well as consistency over time. For international comparative purposes, the IARC has suggested a rather simple set of rules. In brief, these rules state the following:

(1) The recognition of the existence of two or more primary cancers does not depend on time.

(2) A primary cancer is one which originates in a primary site or tissue and is thus neither an extension, a recurrence nor a metastasis.

(3) Only one tumour shall be recognized in an organ or pair of organs or tissue (as defined by the three-digit rubric of the ICD). (This rule may have to be reviewed when ICD-10 comes into effect, for bone, for example, which has been divided between two three-digit rubrics).

(4) Rule 3 does *not* apply if tumours in an organ are of different histology. Table 3 (adapted from Berg, 1982) lists eight major groups of carcinomas and non-carcinomas. The specific histologies (the groups numbered 1, 2, 3, 5, 6 and 7) are considered different for the purpose of defining multiple tumours; groups 4 and 8 include tumours which have not been satisfactorily typed histologically, and cannot therefore be distinguished from the other groups.

The IARC also drew up the following definitions relating to this field:

Multifocal: Discrete, i.e., apparently not in continuity with other primary cancers originating in the same primary site or tissue (e.g., bladder).

Multicentric: Primary cancer originating in different parts of a lymphatic or haematopoietic tissue.

In line with the above rules, both multifocal and multicentric tumours would only be counted once, unless of different histology.

It is strongly recommended that the above definitions should be used when reporting incidence for international compilations such as *Cancer Incidence in Five Continents*. It should be stressed that these simplistic rules may not suffice for clinical studies.

Table 3. Groups of malignant neoplasms considered to be histologically 'different' for the purpose of defining multiple tumours (adapted from Berg, 1982)

	I.	Carcinomas
1		A. Squamous 805–813[a]
2		B. Adenocarcinomas 814, 816, 818–823, 825–855, 857, 894
3		C. Other specific carcinomas 803–804, 815, 817, 824, 856, 858–867
4		D. Unspecified (Carcinomas NOS) 801–802
5	II	Lymphomas 959–974
6	III.	Sarcomas and other soft tissue 868–871, 880–892, 899, 904–905, 912–934, 937, 949–950, 954–958
7	IV.	Other specified (and site-specific) types of cancer 872–879, 893, 895–898, 900–903, 906–911, 935–936, 938–948, 952–953
8	V.	Unspecified types of cancer 800, 999

[a] The numbers refer to the first three digits of the ICD-O morphology code

Coding of neoplasms on death certificates: implications for cancer registries

Most cancer registries have access to death certificates. Ideally a registry should be able to match its records against all deaths, irrespective of stated cause. This so-called "death clearance" enables registries to calculate survival and uncover deaths ascribed to cancer which had not been previously reported to the registry. While many registries have access to all certificates, some obtain information only about those coded to cancer and, unless multiple-cause coding is performed, will learn only about neoplasms considered to be the underlying cause of death. The selection and coding rules for deciding on the underlying cause of death are complex and merit study as their interpretation may influence the coding of neoplasms. The 10th Revision of ICD provides a new rubric for malignant neoplasms of independent (primary) multiple sites (C97), which would normally be used for death certificate coding. In essence this rubric draws attention to the existence of more than one independent primary neoplasm, but does not identify their locations, whereas the coding rules for ICD-9 forced the choice of one site and information on the existence of the other neoplasm(s) was lost. While cancer registries are normally able to identify the existence of multiple independent primary tumours, their handling on death certificates can give rise to problems.

Consultant advice

Information reaching the registry about a given tumour may be incomplete. This may be due to an absence of information or to careless completion of the relevant forms. Rather than guessing, every attempt should be made to contact the notifier who may

be able to provide further information. Nonetheless, all registries should have available a medical consultant who is familiar with the codes used in the registry to help resolve difficult problems. For example, it is often difficult to determine whether a tumour originated in the rectum or colon. If possible, this consultant should review such cases and make the decision. Another difficult site is liver. Whether the registry uses ICD or ICD-O, a decision as to whether a cancer in the liver is primary or secondary may have to be made. If secondary, or unsure whether primary or secondary, the primary site should be coded as being unknown. When ill-defined sites such as arm, leg or other regions of the body are used, the indexing of ICD-O provides help. The histology should indicate what type of tissue the tumour came from: carcinomas are likely to have arisen in the skin, sarcomas in connective tissue and osteo- or chondrosarcomas in bone. If none of these terms is found, then the appropriate ill-defined site, 195.- must be used.

Retrieval and reporting

Coding is of little use if the data cannot be retrieved. Both ICD and ICD-O are well adapted to retrieval. All registries should retrieve and tabulate their data at least annually (for a detailed description see Chapter 10). The very minimum should be a table by site, by sex, and according to the code in use, ICD or ICD-O. If ICD-O is used for coding it should be converted to ICD for tabulation purposes. Only if this is impossible should tabulation by the topographic codes of ICD-O be performed, and these should be supplemented by tables separating the various histological categories. Since there are nearly a thousand histological types, a certain amount of grouping of histologies is necessary. This can be done on a site-by-site basis, listing the common entities. An estimate of likely frequencies can be obtained by consulting Cutler and Young (1975) and Young *et al.* (1981).

In retrieving data over time (trends), it may be necessary to undertake some conversion or regrouping for certain sites. Each ICD revision—7 to 8 to 9 to 10—made certain changes and the user must carefully examine the changes for the site being studied. Not only have code numbers changed, for example, breast has changed from 170 in ICD-7 to 174 in ICD-8 and 9 (for females) and to C50 in ICD-10, but the content of categories has changed as well. For example, in ICD-8 there was only one *in situ* category—that for the cervix uteri (ICD-8 234.0). All other *in situ* neoplasms were counted as malignant neoplasms. A change of codes can be taken care of (see the next section), but the impact of change of content is very difficult to assess.

Tables of ICD conversions

As new classifications and new revisions of ICD have come into use, to report long time series, cancer registries need to convert data coded by previous classifications to the new codes. A registry may maintain its files according to ICD-O but report its results by, say, ICD-9 for annual reports and for inclusion in the series *Cancer Incidence in Five Continents*. The National Cancer Institute in the USA has produced a series of conversion tables for neoplasms, edited by Percy. The recent and current conversions are available on magnetic tape as well as being documented in manuals. Those currently available are for ICD-8 to ICD-9 (Percy, 1983a), ICD-9 to ICD-8

(Percy, 1983b), neoplasms ICD-O to ICD-8 (Percy, 1980), and ICD-O to ICD-9 (Percy & van Holten, 1979).

Many workers have expressed a wish to have conversion from ICD-9 to ICD-O. Data can easily be converted from a detailed to a less detailed version, but not in the other direction. As noted above, most of the terms in the ICD are topographic and the morphology of a malignant tumour is not taken into consideration except for malignant melanoma, choriocarcinoma, the soft tissue neoplasms, the lymphomas and the leukaemias. It is possible to convert the topography but not the morphology. For example 162.9, a malignant tumour of the lung in ICD-9 could be translated into T-162.9 in ICD-O but the morphology field would perforce have to be left blank (—/3) in ICD-O, and an ICD-9 to ICD-O conversion would thus have little value.

In converting from one revision to another, the user should be aware that many terms listed only in the alphabetical index are sometimes indexed differently from one revision to another, and if this term is of considerable frequency it can affect statistics. An example of this is neuroblastoma: this term was indexed, if no site was mentioned, in ICD-8 to 192.5—sympathetic nervous system; in ICD-9, it is indexed to 194.0—adrenal gland. This resulted in a large apparent increase in mortality from adrenal gland cancer when ICD-9 came into use (C. Percy, personal communication).

Since the comparison of incidence data over time is an important function of the cancer registry (see Chapter 3), some registries have chosen to have their cases coded by two different classification systems (e.g., Iceland and Denmark). This is largely facilitated by the extensive use of computers in the registration process. The Danish Cancer Registry's data for the period 1943–1977 are thus coded according to an extended version of ICD-7. Incident cases from 1978 onwards have been coded according to ICD-O and a computer- based conversion table automatically allocates the corresponding ICD-7 code, thus allowing direct tabulation of comparable incidence figures for a period of more than 40 years.

Revisions of ICD

Instead of the usual ten-year period between ICD revisions, it was decided by WHO Member States to lengthen the span for ICD-9 to 15 years since the tenth revision was planned to be a major one.

The periodic revision of ICD raises problems for cancer registries (and for other users and providers of health statistics) in that, unless carefully carried out, it becomes very difficult to compare data over long periods of time. If thought has been given to the problems of time series, it should always be possible to convert from the new revision, usually more detailed, to the previous one, by collapsing information (see also below). Revisions increase the work for all statistical systems, as new computer programs and editing checks have to be written, and output tabulations devised, and registry staff who have learned one set of code numbers have to learn a new code, giving rise to delay and a certain amount of error.

It is of the greatest importance that suggested changes be assessed by field trials before being adopted, as with the prolongation of the period of currency of a revision, mistakes take longer to correct. In this context, the second edition of ICD-O was the subject of extensive field trials.