5
Input Procedures

The preceding chapters describe finding information on cancer cases in the registration area, recognizing and abstracting the relevant data and putting the information about the cancer diagnoses into coded form. This section looks at how the different documents which come to the registry can be managed, so that records relating to a patient from more than one source are linked to the same patient, ensuring that all the information necessary is present and avoiding duplicate registrations. Once documentation on each case is complete the data have also to be stored in such a way that they are available for consultation, updating and analysis.

The principles of the input operation are standard for all registries, although methods of processing the data will differ. Notifications have to be checked for completeness and accuracy, linked to the registry database so that it is possible to identify new cases versus new information on cases already registered, coded and added to the database or filed elsewhere. If the registry is computerized the process is simplified, but the steps to be followed are similar in both manual and computerized registries. It is assumed here that the majority of cancer registries are now working with computers, but reference is made to manual operations. Many computerized registries maintain some manual operations within a largely automated system.

Figures 1 and 2 provide illustrated examples of the flow of operations in two different registries, the first illustrating the treatment of hospital abstracts in the Rizal (Philippines) Cancer Registry and the next the procedures carried out for all notifications in the Danish Cancer Registry.

5.1 Receipt of notifications

Notifications on cancer patients will come in different formats, at different times and in different ways. Some registries receive routine notifications from hospitals, pathology laboratories and death records on a regular basis. In others the registry clerk(s) go to the various data sources to fill in the registry abstract forms. The registry may receive abstracts or notifications daily or in batches. The documents can be processed at varying intervals; for larger registries, for example, they may be dealt with in weekly batches (see Figure 2, which describes the processing of some 1000 forms weekly), and for smaller registries processing of data may be done by hospital and by year.

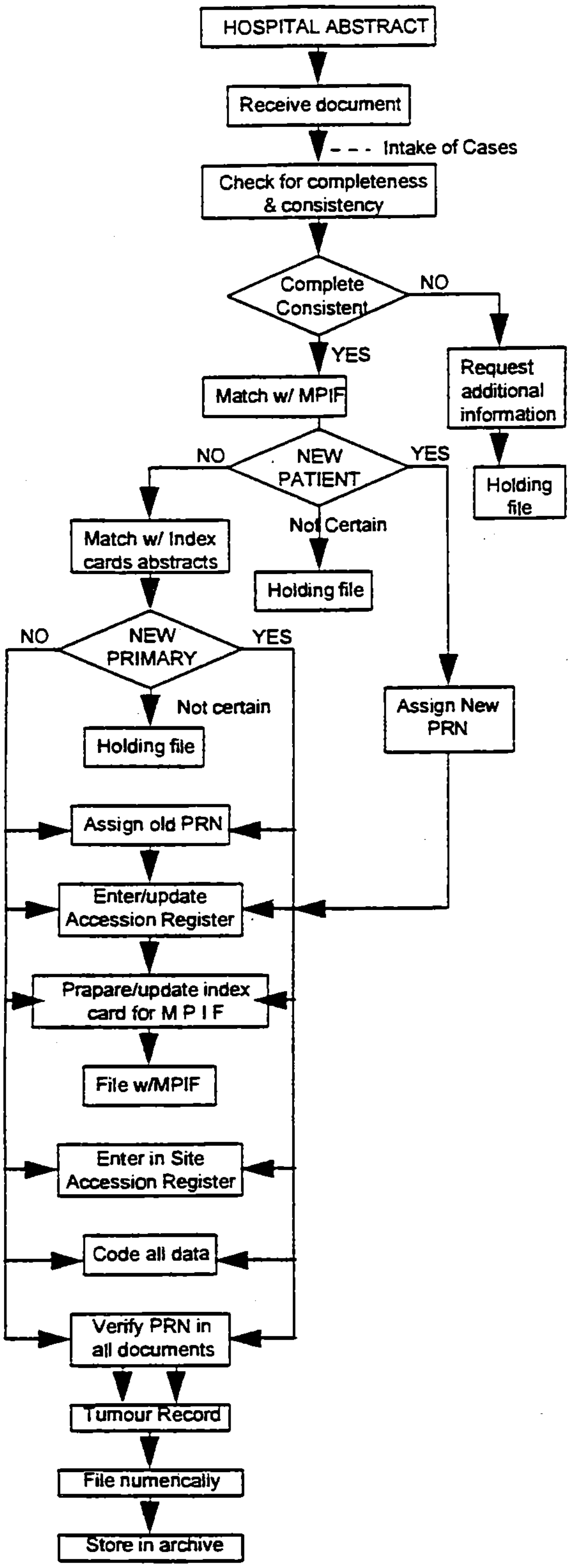
It is useful to maintain a record of the cases abstracted from each hospital in the area by year, to provide a summary record of the number of cases abstracted by hospital or hospital department, the distribution of cancer cases in each hospital by site and the workload of the registry personnel in the hospitals. In a computerized database this type of list is generated by using 'Source of information' to sort the records.

If the registry has access to death certificates, they may be sent automatically to the registry or it may be necessary to go to consult them in the government office responsible. When death certificate notifications are received it may be advisable to allow time for any hospital abstracts to be completed before undertaking the follow-back process (see below).

The registry should also establish a list of reasons why a notification should not be entered into the system in the standard way, such as:

- address not within the defined registration area: these cases should be manually filed separately or labelled as 'non-resident' in a computerized database. If they are resident in an area covered by another population-based registry the details should be transferred to that registry;

Figure 1. In-put procedures for hospital abstracts (Rizal Cancer Registry)



1. Receipt of Document

   Stamp with date when document received

   Check for completeness and consistency

   Suspense (Holding) File for incomplete or inconsistent abstracts

2. Record Linkage

   Match with:

   Master Patient Index File (MPIF)

   File of Prior to Reference Date Cases (FPRDC)

   Identification of Old and New Patients

   Identification of Old and New Primary Tumours

3. Registration of New Patients

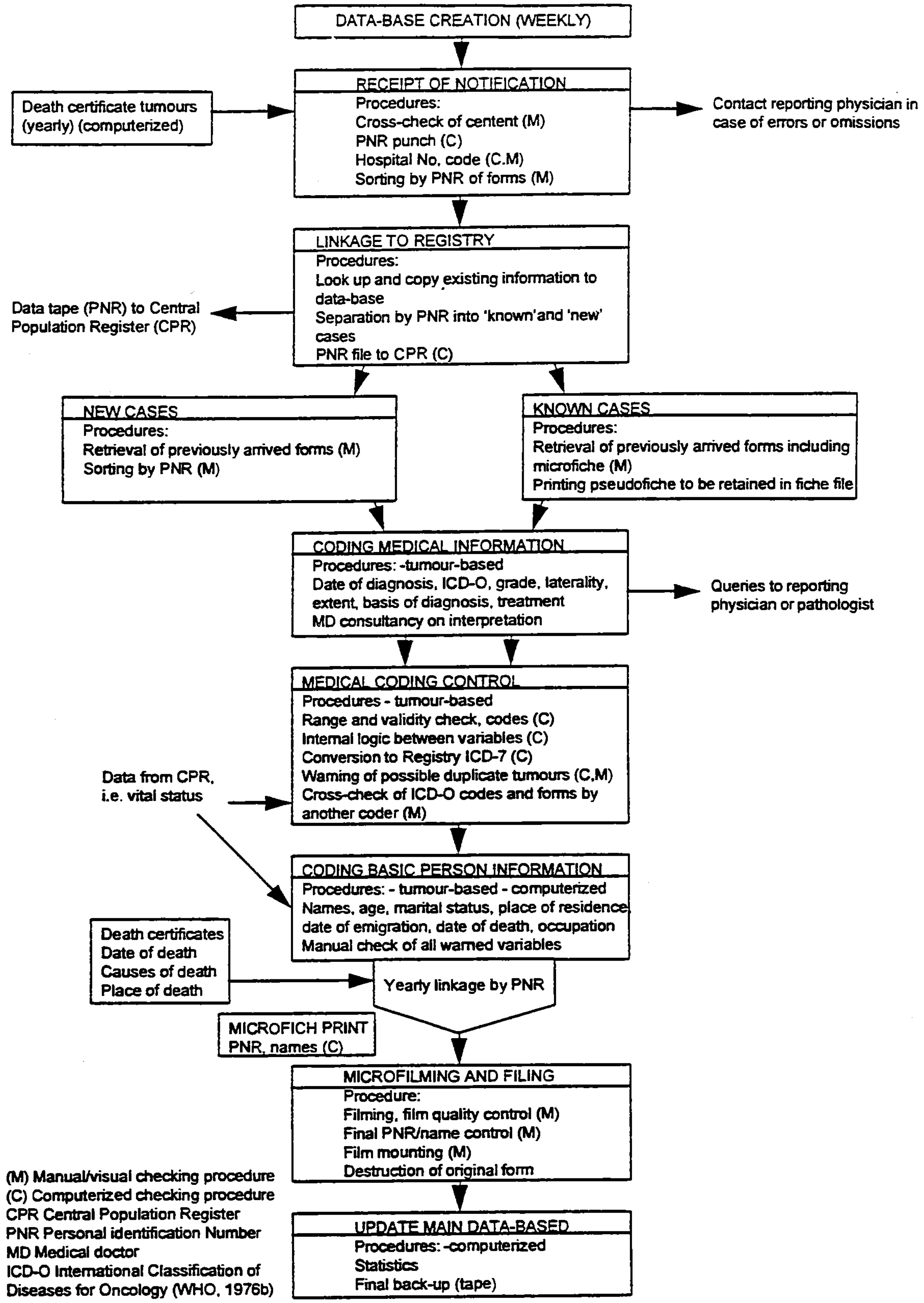
4. Registration of New Primary Tumours

5. Updating of Old Tumour Record

6. Filing the Tumour Records numerically, based on PRN (patient registry number)

7. Storage in archives

Figure 2. Flow diagram (Danish Cancer Registry)



- incidence date occurring earlier than the registry's reference date (the date from which the registry started registering cancer cases). These cases should be filed apart or labelled in the computer, and while they do not appear in the incidence data the information can be very useful, making it possible to check whether or not a case was diagnosed before the time the registry began operations. This is particularly true for registries which are just starting.

5.2 Checking for completeness and consistency

The validity of the data in the cancer registry should be looked at throughout the registration process. Chapter 6 describes some of the measures which can be used to assess the quality of the data in a cancer registry, both in relation to completeness of coverage (whether the registry is obtaining information on all the cases of cancer diagnosed in the registration area) and to the accuracy of the data.

The first stage in this continuing process of quality control is a visual check of the notification forms for content, and this should be done soon after receipt of any documents. If information necessary to register a case is missing, or if entries on the form are not logical, the information necessary to complete or clarify the form should be requested or looked for.

When information is being abstracted from hospital records or from any other source, if any data item is missing this should be noted on the form, e.g. NS (not specified) in the space for date of birth when date of birth is not given in the original records.

Computerized data checking may be done on-line (at the time of data entry) or as part of a batch operation. In on-line systems the data are put directly into the computer and errors or inconsistencies are automatically signalled by the system. In a batch operation the system will produce a listing of errors and queries on cases. Whether the data are checked on-line or in batch mode, the system should signal inconsistencies such as age not corresponding to date of birth, sex-specific tumours occurring in the wrong sex (prostate cancer in a woman), or date of death before incidence date or date of birth.

The degree of checking for inconsistencies and errors, as well as the extent to which a registry pursues missing information, will influence the quality of the data.

5.3 Missing or incorrect data

Abstracts on which essential data are missing or wrong should be kept in a temporary or holding file until the necessary information is obtained. In a manual registry this would be a physical file, but on a computer database such cases should be entered and labelled as 'suspense' cases.

Cases for which information is missing or notifications which need correcting should be followed back to the hospital sources so that the abstract form can be properly filled in. In the meantime the record can be labelled as a 'suspense' case or the abstract kept in the holding file. Similarly, for cases where the diagnosis remains doubtful, the details of the case should be abstracted but labelled 'suspense' or kept in the holding file.

5. 4 Coding of data

To enter information on paper into the computer, it is necessary to code a number of the data items. The topography and morphology are coded according to a recognized international classification, generally ICD-O (see Chapter 4). Additional dictionaries of codes can be created for many of the data items collected in a registry, such as basis of diagnosis, stage, ethnic group, hospital, area of residence. It is important to remember, when creating dictionaries, that an 'unknown' category is always required, e.g., for ethnic group:

1	Chinese
2	Malay
3	Indian
4	Other
9	Unknown

A computer can do the coding automatically, either working from a written description to find and allocate the appropriate code, or producing a written description when the code has been entered.

5.5 In-put

5.5.1 Linkage

The registry receives records relating to an individual patient from a variety of sources, such as different hospitals, pathology laboratories or offices of vital statistics. These records have to be linked together to complete all the information on each patient and to determine if the case (or cancer) is already registered or not. Remember that the registry is recording cancers, not individuals. If one person has two or three different cancers (according to the rules for defining multiple primaries given in Chapter 4) then two or three registrations are made – one for each cancer. It is, however, essential to avoid duplicate registrations, i.e. registering the same cancer more than once, and this is why this process of linkage of different documents is so important.

In countries where a unique personal identity number is used the linkage process is greatly facilitated. However, in countries which do not have such a number (the majority), matching the patient's name with the registry files is usually the only method. For names which are very common, the name alone is not sufficient for identification and other items such as date of birth and address are also necessary. When matching records, it must be borne in mind that data items may not be consistent between different documents. Names may be misspelled due to mishearing or to miscopying, or a patient may use different names in different institutions or on different admissions to a hospital. For example, a married woman may use her maiden name on one admission and her married name on another, so to the registry it appears that there are two diagnoses and a duplicate registration results. All known names for each patient should be recorded.

The information considered essential for registering a case is:

- name and address (or equivalent minimum identifying information, depending upon local circumstances);
- age or date of birth;
- sex;
- site;
- morphology (histology);
- basis of diagnosis;

Each document coming into the registry must be checked to see if it relates to a person already registered.

Manual linkage

In a manually operated registry incoming documents are checked with an index, generally on cards and arranged alphabetically by name. This file should include both living and dead cases, and each index card would contain information on the full name of the patient, the registration number, age, sex, address and other details such as source of information, hospital case number, incidence date, primary site, morphology and (if applicable) date of death. When matching by name, allowance must be made for errors in spelling (phonetic spelling of names or errors due to illegibility of hospital records). If there is a match in the name, then the age, sex, address and diagnosis are compared to make sure that the new information belongs to the previously registered patient.

If there is no match, then this is a new patient who should be registered. If it is uncertain whether the case is a new patient or not, the case may have to be followed back to the data source for additional information and in this instance the notification should be held in the holding file.

On-line record linkage

Some computer systems have sufficient flexibility to permit on-line searching for possible matches using identifying variables (for example name, date of birth and address). The principle is the same as in a manual registry, but instead of the index being stored on cards it is held on disk and the software takes the place of the human searcher. The index file is maintained by the computer, and when the operator types in the name, date of birth, address and other identifying information the computer searches its files for cases with the same or similar details and displays possible matches on the screen. Based on this, the operator decides whether the case is known to the registry (patient already registered) or is a new patient. Such systems may have elaborate facilities for identifying possible matches using phonetic methods. As soon as a new patient is identified, the computer automatically sets up a new record. Changes can be made and additional information added to this record at any time. Incoming documents do not have to be arranged alphabetically or in any particular order for the record linkage.

Off-line (batch) record linkage

Off-line record linkage is done by creating a batch of completed records on disk or tape and comparing this with the existing records. A scoring system may be used to establish the degree of matching of each new record, so that an exact match achieves a higher score than a close or similar match, and the absence of a match results in a zero score. A weighted score is then computed and the score evaluated. Above a fixed level the match is presumed to be correct, and below another fixed level the absence of a match is assumed. Even with sophisticated computer systems this stage will probably be done by registry personnel. The computer produces a listing of possible matches, but the final judgement is left to the registrar. Matched records automatically update the existing records and unmatched records set up new registrations.

5.5.2 Registration of a new patient

When a new patient is identified (record linkage has shown no existing record for the patient), a registration number (also called accession number) is assigned and a new record created. This is done automatically in a computerized system. The number is a unique number which is used to identify the patient, and its composition will vary from one registry to another. The most common example is the use of the first two digits for year of diagnosis, followed by a number allocated serially as new cases are registered.

5.5.3 Identification of a new primary tumour

When there are several records for the same patient, the most appropriate primary site, histology and incidence date have to be determined. The site is selected from the most reliable source, i.e. a pathology report will be preferred to a clinical diagnosis. If more than one three-digit site code remains, the histology codes on each record are compared. If the histological diagnosis is different (see Chapter 4 for definitions of multiple primaries), a second primary tumour is identified and a new registration made.

In a manual system lists of all the cancer cases registered each year should be kept, one arranged by order of registration and one by site. The site list can help to check for duplicate registrations since cases with the same diagnosis will be grouped.

5.5.4 Second primary cancers

If a new primary tumour is identified in an individual who is already registered (multiple tumours), a new registration has to be created. In a sophisticated computerized system the same registration number can be used for one individual with two or more tumours, and the tumours are identified by a tumour number. In most systems, a new registration has to be created and it is advisable to have a separate field in which it is possible to identify multiple cancers.

5.5.5 Updating the records

Once the record linkage process has been carried out, new information linked to a case already registered should be used to update the existing record. The type of information which may be received includes:

- more precise identifying information such as middle or maiden name, date or place of birth, sex, civil status or a more complete address;
- a more accurate incidence date, a more valid basis of diagnosis, or a more definite primary site or histological type (e.g. as provided by autopsy findings);
- for patients who have died, the date, place and cause of death.

5.5.6 Death certificate notifications

The information on the death certificate is linked in the same way as that from a hospital abstract. If the case has been registered previously the record is updated with the date of death. If there is no registration for this case, the place of death is checked. If the patient died in a hospital, the case is followed back to the hospital and a hospital abstract is prepared. If the patient died at home, the case-finding lists from the different hospitals (see Chapter 3) may be used to check whether or not the patient was seen in a hospital. If yes, the case is followed back to the hospital.

If there has been no previous registration, and it proves impossible to trace any record of the case having been seen in hospital or by a physician, the case is registered as a new cancer and the Basis of diagnosis is 'Death Certificate Only' (DCO), while the Source of information will be 'death certificate'. Figure 3 illustrates the process of registering a death certificate notification.

5.6 Record maintenance

5.6.1 Editing

All documents are checked for completeness and consistency as soon as they reach the registry. However, further omissions or inaccuracies may be found later, either due to errors in data entry or coding or because additional information has been received. The data in the cancer registry can be edited (added to or corrected) before, during and after entry.

5.6.2 Retrospective checking for duplicates

In spite of repeated checks, some duplication of registration may occur in the following circumstances:

- the same patient may use different names, e.g. maiden name in one hospital and married name in another;
- some patients use nicknames or aliases;
- errors in the spelling of the patient's name may not have been noticed during the linkage process;
- inaccurate information on, for example, the date or place of birth may have led to registration of two patients when there was only one.

When duplicate registrations are discovered, all existing records must be drawn together to update one record and cancel the other(s).

5.7 Queries and follow-up

5.7.1 Looking for information

It is often necessary for the registry to make enquiries about missing data or information which is obviously inconsistent. In the manual registry, either standard letters giving the known details about the patient and describing the problem are sent out or the registry clerk visits the hospital(s) to try and find the information. The same procedure is used in a computerized registry, but it is possible to develop a system which will detect missing or inconsistent data and generate a query automatically.

5.7.2 Follow-up

If the registry practices active follow-up enquiries are made at regular intervals, usually once a year on the anniversary date of the patient's diagnosis, about each patient not known to be dead. In a manual system, index cards for all patients to be followed-up should be kept grouped by the month when follow-up is due. When the date of follow-up arrives a form is sent to the hospital or general practitioner treating the patient. If a notification of death is received the card concerning the individual should be removed from the follow-up files.

The procedure is the same in a computerized system, but the computer can automatically generate the follow-up requests.

A passive follow-up system relies on receiving notifications of deaths of registered cases, and it is not necessary to generate routine enquiries.

5.8 Storage

Documents on paper must be filed using the registry filing system (usually numerically by patient accession number) and stored in a room which is secure and inaccessible to unauthorized persons. They should be protected against loss or damage from fire, floods or interference. If index cards are used they should be kept in filing cabinets for index cards.

Computerized data are stored on magnetic tapes or diskette. The cancer registry database exists on hard disk, so that it is possible to process records in any order, and working on-line, to alter one record without disturbing the others. Records can be accessed randomly or in various index sequences (numerical, alphabetical, by primary site, etc.). Several software packages exist to maintain these databases, but some training and experience is needed to be able to use them

Magnetic tape files consist of a series of records, with each cancer usually occupying one record. The records are read in the order they are held on the tape. There is no limit to the size of the file and magnetic tape is an economic and efficient way of backing up the cancer registry data, so ideal for storage.

Diskettes In a registry which only has access to Pcs, the data should be backed up on diskette. Data held on diskette have the same advantages as those on the disk – they are easy to work with, and records can be accessed in the same way. The disadvantage when compared with magnetic tape is that they hold much less data.

Computer files, both programs and data, should always be kept in at least two different places to guard against damage or theft. Back-up (on tape or diskette) of data should be done regularly so that data can be recovered in cases of hardware failure or accidental deletion. If there are a lot of records a full back-up can be done at regular intervals, say at the end of each month, and an incremental back-up of just the records entered or amended since the last full back-up can then be done each day. If all the data are lost, they can be restored using first the last full back-up and then the incremental back-up.

Figure 3. Flow chart for death notifications