

Appendix 4. CANREG: Cancer registration software for microcomputers

M. P. Coleman and C. A. Bieber

*International Agency for Research on Cancer,
150 cours Albert Thomas, 69372 Lyon Cédex 08, France*

Introduction

The use of computers to store, manipulate and analyse cancer registration data is now widespread, and the range of hardware and software in use is considerable. In a recent survey on the use of computers in 61 cancer registries of all sizes and from all regions of the world, there were 20 different hardware manufacturers, 12 different types of data-base management software, and 30 different computing languages and statistical packages in use (Menck, 1986). Only 24 of the registries used microcomputers, and in only five was a microcomputer the primary or only computer in use.

In many small or recently started cancer registries, however, and in particular those in developing countries, microcomputers are the only feasible option for computerized data entry and manipulation. Microcomputers are cheap and simple to operate, they require little maintenance and (in most cases) are mutually compatible, and they provide access to a large range of commercial software for other registry functions, e.g., text processing.

This appendix provides a brief description of CANREG, a software package for cancer registration using microcomputers. CANREG was designed specifically for cancer registries with very limited budgets and with little or no computing expertise, and for use with relatively small data sets: a maximum of 45 items of data per tumour and up to about 3000 cases per year. It was designed to be usable by registry personnel who may have very limited education—perhaps only six to ten years of schooling. CANREG is not suitable for mini or mainframe computers, for which more powerful software packages are available. Large registries in the USA may spend US \$500 000 or more on developing and testing their software: CANREG can be installed on systems costing as little as \$6000.

Main features of CANREG

CANREG enables personal, demographic and tumour-related data about individuals with cancer to be entered into the computer, checked automatically for data completeness and validity, examined, corrected and finally used to prepare listings, frequency tables and cancer incidence statistics. The program is menu-driven, and

Data Entry Fiji Cancer Registry		25-01-1989
Registration No	844282	
Site	1429	Major salivary gland
Hospital	2	Lautoka
Multiple Primaries	2	No
Surname	DOE	First Name JOHN Other Names ...
Domicile	1001	Namosi
Date of Birth	311234	
Age	49	
Sex	1	Male
Race	1	Fijian
Diagnosis Date	200384	
Histology	81403	Adenocarcinoma, NOS
Basis	7	Histology of Primary
Date of Death	
Post Mortem	.	

Figure 1. Example of CANREG data entry screen from the Fiji Cancer Registry
 Note: the names and data shown are fictitious

can be operated by personnel with very little training. The system is user-friendly and flexible, and can be repeatedly modified to the needs of a developing registry.

Data entry: individual cases

Data for each case of cancer are usually entered at the keyboard, using a data entry screen designed to meet each registry's requirements (see Figure 1).

Many of the data items are entered in coded form and automatically checked against dictionaries in which all possible valid entries for each data item are stored (e.g., 1 = male, 2 = female, 9 = unknown). At the moment of data entry, the meaning of the code is displayed on the screen, enabling the operator to confirm that the original data have been correctly coded: thus an entry of 80503 for morphology would result in 'Papillary carcinoma, NOS' being displayed, for comparison with the written record. Codes not included in the dictionary for that item would be rejected.

Up to 30 variables may be encoded in dictionaries in this way, and the CANREG system is supplied with dictionaries for the topography and morphology codes of the International Classification of Diseases for Oncology (ICD-O) either in English or French. All dictionaries can be easily modified, and new dictionaries created. Thus the ethnic groups in the registry's population can be included in a dictionary with the name of each group in the local language. Equally, all hospitals, laboratories and other sources of data can be encoded in a separate dictionary.

A simple data entry screen for an imaginary patient is shown in Figure 1. The data entry clerk has typed 1429 for site, and the corresponding topography is displayed next to this code. The hospital has also been entered as a code (2), and the name of the corresponding hospital has been displayed from the appropriate dictionary. Dictionary control of data entry is an important method of maintaining the internal validity of the registry data.

Table 1. CANREG internal logic checks^a at data entry

-
- 1 Registration number has not already been assigned
 - 2 Sex and site are compatible
 - 3 Dates are valid, possible calendar dates
 - 4 Dates are mutually compatible
 - 4.1 Date of birth earlier than dates of diagnosis and death
 - 4.2 Date of diagnosis not later than date of death
 - 4.3 Year of diagnosis and death not later than current year
 - 4.4 Age consistent with dates of birth and diagnosis
 - 5 No missing information for essential variables
 - 6 Codes for dictionary-encoded variables are valid
 - 7 Numeric variables do not contain alpha codes
 - 8 No embedded spaces (optional)
-

^a The variables on which checks 5 to 8 are carried out can be defined by the user.

Certain variables are essential in every registration record, such as the patient's name and date of birth, and the tumour site. The program ensures that all such variables have been supplied before allowing a new record to be added to the data-base. The choice of which variables are to be treated as essential can be defined and later modified by the user.

The internal logic of each record is also subjected to a number of checks at the moment of data entry (see Table 1). Records which fail any check are not added to the data-base, and a message inviting a suitable correction is displayed.

These checks on the data entry procedures are designed to improve the quality of the registry data by ensuring that every tumour record added to it conforms to a consistent standard of completeness and validity.

Use of coded data

The deliberate choice to enter and store most of the data in CANREG in coded form was made for several reasons.

Storage efficiency

In most cases it requires less space to store a code than the corresponding text which it represents. For data items such as morphology, the economy in storage space from using codes may be considerable (e.g., 5 digits instead of 44 letters).

Speed of entry

Entry of codes is quicker and less prone to error than entry of text. Codes do need to be assigned manually prior to data entry, but then coding the data is always an important stage of checking the validity of the record.

Validity checks

These are easier to arrange if the data are numerical.

Language compatibility

Use of codes for topography and morphology enables data produced in, say, an English-language version of CANREG to be interpreted and analysed without modification on a French-language version, since the dictionaries are numerically equivalent, with the exception of five or six rarely used rubrics.

Data entry: importing files

Data files already on computer can be imported into a CANREG data-base if they are in a standard (ASCII) format. The sequence of variables in the data records does not need to be the same as in the CANREG data-base. The logic checks performed when entering individual records, however, are not carried out on imported data sets.

Data management

Tumour records in the data-base can be called up for display on the screen by entering the registration number; records can then be updated or corrected, e.g., with the date of death or a revised histology report. The internal logic checks described above are also performed when a record is altered, and before the data-base is updated. Records can also be deleted at this stage.

Data security

The program incorporates password protection of the entire data-base. The program itself is provided in compiled form, and cannot be readily examined in order to decipher the data-base. An additional advantage of encoding many of the variables is that data records are largely numerical. If a diskette containing the master data file were to be lost in the mail, it would be uninterpretable without all the associated dictionaries and the files which define the record structure.

The program incorporates a facility for making back-up copies of the data-base on diskette, and for restoring the data-base from the back-up diskette in the event of failure of the hard disk used as primary storage. Users are recommended to perform back-ups at weekly intervals, or more often if large volumes of data are being entered.

Subsets, listings and tables

The CANREG system allows subsets of the data to be selected, using up to three variables at a time (e.g. site, sex, age), in order to produce files containing, say, all lung cancers in males aged 40–69. The selection procedure can be repeated on the original subset, using different variables, to produce a more specific subset defined additionally by, say, histology.

Listings of individual cases can be produced, either printed or displayed on the screen, in a wide variety of styles which can be defined and modified by the user. Listings can also be produced as standard (ASCII) files, for manipulation by other programs. Listings can be sequenced (e.g. by site, sex and name) and restricted to a

Page 1	Patient Name List (with Sex/Age/Site/Hist)						25-01-1989
Registry Number	Surname	First Name	Age	Sex	Site	Site Description	Hist Histological Description
840096	MONEA	VALERIANO	54	Mal	141.0	Base of tongue, NOS	8070/3 Squamous cell carcinoma, NOS
850101	OZNUMBA	GEORGES	60	Mal	141.0	Base of tongue, NOS	8070/3 Squamous cell carcinoma, NOS
840030	UOMIRI	EMILE	60	Mal	141.9	Tongue, NOS	8070/3 Squamous cell carcinoma, NOS
850045	NAMFOUBI	JOSEPH	53	Mal	141.9	Tongue, NOS	8070/3 Squamous cell carcinoma, NOS
840242	REVDANT	LOUIS	58	Mal	141.9	Tongue, NOS	8070/3 Squamous cell carcinoma, NOS
860194	DABIAGI	VERONIQUE	54	Fem	142.0	Parotid gland	8020/3 Carc., undifferentiated type, NOS
850060	BAMICKA	DELPHINE	32	Fem	142.1	Submandibular gland	8200/3 Adenoid cystic carcinoma, NOS
840035	KOMAGNI	WILFRIED	10	Mal	143.1	Lower gum	9750/3 Burkitt's tumor
850116	NANG BEKALE	JOSEPH	14	Mal	143.1	Lower gum	9750/3 Burkitt's tumor
840031	NAMFOUMBI	JOSEPH	65	Mal	143.9	Gum, NOS	8070/3 Squamous cell carcinoma, NOS
850065	NOMGOKO	JEAN FIDEL	09	Mal	143.9	Gum, NOS	9750/3 Burkitt's tumor
840230	FASOU PEFOUKA	CLAIRE	57	Fem	143.9	Gum, NOS	8070/3 Squamous cell carcinoma, NOS
840019	UOHBAGA	MATHIAS	69	Mal	144.8	Other	8070/3 Squamous cell carcinoma, NOS
850232	BASI	HILAIRE	47	Mal	145.1	Vestibule of mouth	8070/3 Squamous cell carcinoma, NOS
850227	TEROA	GEORGES	25	Mal	145.2	Hard palate	8051/3 Verrucous carcinoma, NOS
850034	UOMSSAVOU KOUNBA	ADOLPHE	67	Mal	145.3	Soft palate, NOS	8070/3 Squamous cell carcinoma, NOS
850064	OBHUMBA	EDOUARD	33	Mal	145.3	Soft palate, NOS	8070/3 Squamous cell carcinoma, NOS
840067	BOHILE	ROBERT	54	Mal	145.3	Soft palate, NOS	8070/3 Squamous cell carcinoma, NOS
850191	GNUULET	HENRI	78	Mal	145.5	Palate, NOS	8230/3 Solid carcinoma, NOS
860193	SIBSINE	MADELEINE	32	Fem	146.0	Tonsil, NOS	8070/3 Squamous cell carcinoma, NOS
840225	NORAR	GASTON	44	Mal	146.0	Tonsil, NOS	9692/3 M. l., cnttblst-cntrcyt, foll
840237	MOTBE	PAULINE	21	Fem	146.0	Tonsil, NOS	9630/3 M. l., lymphocyt, poorly diff., NOS
840017	ABM	MATHIAS	40	Mal	148.1	Pyriiform sinus	8070/3 Squamous cell carcinoma, NOS
850062	KIMALA	JEROME	60	Mal	148.1	Pyriiform sinus	8070/3 Squamous cell carcinoma, NOS
840229	NOSA	JEAN MARIE	50	Mal	150.9	Esophagus, NOS	8070/3 Squamous cell carcinoma, NOS
840222	UOBENGA	ANDRE	31	Mal	151.2	Pyloic antrum	8140/3 Adenocarcinoma, NOS
850141	KELOULEPEGUE	JACQUES	Mal	151.9	Stomach, NOS	8890/3 Leiomyosarcoma, NOS	
840156	OSTNGA	NORBERT	50	Mal	151.9	Stomach, NOS	8140/3 Adenocarcinoma, NOS
850217	EFN	FLAVIENNE	70	Fem	151.9	Stomach, NOS	8140/3 Adenocarcinoma, NOS
840218	NIMDOUGOU	JOSEPH	57	Mal	151.9	Stomach, NOS	9140/3 Kaposi's sarcoma
850219	EKLEYAGA	ODETTE	50	Fem	151.9	Stomach, NOS	9611/3 Mal. lymphoma, lymphoplasmacytoid
840152	TIPAKE	THERESE	60	Fem	153.3	Sigmoid colon	8140/3 Adenocarcinoma, NOS
840012	IAMGA	CLEMENCE	36	Fem	153.6	Ascending colon	8480/3 Mucinous adenocarcinoma
840206	UOBANGA	THEOPHILE	51	Mal	153.6	Ascending colon	8144/3 Adenocarcinoma, intestinal type
850246	KOWO	MATHIAS	60	Mal	153.6	Ascending colon	8144/3 Adenocarcinoma, intestinal type

Figure 2. Example of listing generated with CANREG

Note: the names and data shown are fictitious

particular type of record (e.g. males only) or to a range of values (e.g. site codes 150–159).

The example in Figure 2 shows a listing of (imaginary) cases sequenced by site; in this example, site and histology codes have been printed along with their corresponding text, extracted automatically from the code dictionaries at the time of printing. The example chosen is a listing frequently used by registries of moderate size: it could have been sorted additionally by sex and by name, as one method of searching for duplicate registrations.

Data analysis

Most registries need to produce standard tables on a regular basis, such as annual tables of the numbers of cases registered by sex, age, and site, or similar tables of age-specific incidence rates.

A set of analysis programs is therefore provided with CANREG. The programs are separate from the main CANREG system described above, so that they can be used with data produced from any computer system, but they are designed to accept data in a form which can be readily produced by CANREG as an 'export' file (see below).

The main analysis program (CRGTable) produces tables of the numbers of cases by sex (males, females and all persons), age-group and cancer site or type. The age-groups can be chosen by the user, and the cancer sites or types would be those entered with the data in CANREG. The tables include totals by age (including a category for cases with unknown age) and by site, and the frequency (percentage) of each cancer relative to all cancers. If population data are supplied, this program also produces tables of incidence rates specific for age, sex and site, and crude and age-standardized

* - Female - *		1986 - 1987																Data File --> Sample!.dat	
		Incidence Rate by age group (per 100,000)																	
Site (ICD-9)	TOTAL	0-	5-	10-	15-	20-	25-	30-	35-	40-	45-	50-	55-	60-	65-	70+	Crude	ASR	
ALL SITES	507	5.0	6.1	10.3	20.6	36.9	63.9	155.2	146.5	254.3	248.7	315.4	467.6	322.1	547.8	152.7	77.3	130.2	
Lip	3	-	-	-	-	-	1.9	-	-	-	-	6.3	-	-	16.6	-	0.5	1.0	
Tongue	4	-	-	-	-	-	3.9	-	-	-	5.4	-	-	-	-	7.6	0.6	0.9	
Salivary gland	7	0.8	-	1.3	1.4	1.8	-	-	-	-	5.4	-	9.4	8.9	-	-	1.1	1.5	
Mouth	5	-	-	-	-	1.8	3.9	-	2.9	-	-	6.3	-	-	-	-	0.8	0.9	
Nasopharynx	1	-	-	-	-	1.8	-	-	-	-	-	-	-	-	-	-	0.2	0.1	
Oesophagus	4	-	-	-	-	-	-	-	-	-	-	6.3	9.4	8.9	-	7.6	0.6	1.4	
Stomach	49	-	-	-	2.7	-	1.9	7.1	14.7	49.3	16.2	31.5	56.1	35.8	83.0	15.3	7.5	14.0	
Small intestine	1	-	-	-	-	-	-	-	-	-	-	6.3	-	-	-	-	0.2	0.3	
Colon	2	-	-	-	-	-	-	2.4	-	-	5.4	-	-	-	-	-	0.3	0.5	
Rectum	10	-	-	-	-	3.9	4.7	5.9	3.8	10.8	-	-	-	8.9	-	-	1.5	2.2	
Liver	51	-	1.0	-	2.7	-	3.9	18.8	23.4	26.6	16.2	37.9	18.7	80.5	16.6	15.3	7.8	12.7	
Nose, sinuses etc.	4	-	-	-	1.4	-	3.9	-	-	-	-	9.4	-	-	-	-	0.6	0.8	
Bronchus, lung	4	-	-	-	-	-	-	7.1	-	-	-	6.3	9.4	8.9	-	-	0.8	1.1	
Bone	14	-	-	-	1.4	3.5	1.9	4.7	2.9	7.6	5.4	6.3	9.4	8.9	16.6	-	2.1	3.3	
Connective tissue	7	1.7	-	-	-	1.8	-	-	2.9	-	-	-	18.7	-	16.6	-	1.1	1.8	
Melanoma of Skin	10	-	-	-	-	-	-	4.7	-	-	-	-	37.4	8.9	33.2	7.6	1.5	3.4	
Other Skin	15	0.8	-	-	2.7	1.8	1.9	4.7	-	-	-	10.8	6.3	9.4	-	49.8	7.6	2.3	4.1
Female breast	54	-	-	2.6	-	-	7.7	4.7	14.7	30.4	16.2	44.2	93.5	35.8	83.0	30.5	8.2	15.9	
Uterus	20	-	-	-	-	1.8	-	2.4	-	22.8	5.4	25.2	18.7	-	33.2	22.9	3.1	5.9	
Cervix uteri	102	-	-	-	2.7	3.5	7.7	61.2	38.1	79.7	54.1	56.8	74.8	44.7	16.6	7.6	15.6	23.6	
Corpus uteri	13	-	-	-	-	1.9	2.4	5.9	3.8	27.0	12.6	-	8.9	-	-	-	2.0	3.5	
Ovary etc.	13	-	1.0	1.3	1.4	-	-	2.4	8.8	7.6	10.8	-	18.7	-	-	-	2.0	2.9	
Other female genital	10	-	-	-	-	1.8	1.9	4.7	-	5.4	-	28.1	8.9	16.6	-	-	1.5	2.9	
Bladder	43	-	-	-	1.4	3.5	5.8	14.1	14.7	7.6	27.0	50.5	28.1	26.8	83.0	-	6.6	11.9	
Kidney	6	-	1.0	-	-	-	1.9	-	8.8	-	-	-	8.9	-	-	-	0.9	1.1	
Eye	10	0.8	1.0	2.6	-	3.5	1.9	7.1	-	-	-	-	-	-	-	-	1.5	1.3	
Brain, nervous system	1	-	-	-	-	-	-	2.4	-	-	-	-	-	-	-	-	0.2	0.1	
Thyroid	3	-	-	-	-	1.8	1.9	-	-	-	-	-	-	-	16.6	-	0.5	0.8	
Other endocrine	1	-	-	-	-	-	-	-	-	3.8	-	-	-	-	-	-	0.2	0.2	
Lymphosarcoma, etc	3	-	2.0	-	-	-	-	-	-	-	5.4	-	-	-	-	-	0.5	0.5	
Hodgkin's Disease	6	-	-	1.3	-	7.0	-	-	-	-	-	-	-	-	16.6	-	0.9	0.9	
Leukaemia, cell unspec.	3	-	-	-	1.4	-	1.9	-	-	-	5.4	-	-	-	-	-	0.5	0.6	
Primary Site Uncertain	27	0.8	-	1.3	1.4	1.8	3.9	-	2.9	11.4	16.2	12.6	28.1	17.9	49.8	30.5	4.1	7.8	

Figure 3. Example of a tabulation generated with CANREG's analysis program (CRGTable)

(world) rates (see Figure 3). For childhood tumours, the site groups used are derived automatically from ICD-O site and histology codes (Birch & Marsden, 1987), and the age groups are 0, 1-4, 5-9 and 10-14 years.

Other functions of CANREG

The CANREG system incorporates other programs which provide several additional functions related to cancer registration, such as:

- enabling the operator to check the validity of individual ICD codes and providing the text definition of each code in English, French (ICD-9 and ICD-O) or Spanish (ICD-O only);
- converting ICD-O codes in a file to their corresponding ICD-9 values;
- creating childhood tumour type codes (Birch & Marsden, 1987) in a data file, using the ICD-O site and histology codes, and producing tables of numbers of cases or rates by age, sex and tumour type.

Exporting data sets

CANREG data-bases can be produced in standard (ASCII) computer file format for 'export', either to other centres, for collaborative studies, or for local analysis, using the CANREG analysis programs or other software.

Hardware requirements

The essential requirements to be able to operate CANREG are:

- a microcomputer, including screen and keyboard, which is 100% compatible with an IBM PC, and has at least 256 kilobytes of random access memory (256 k RAM), 10 megabytes of hard disk storage, and a diskette drive;

- an operating system (version 2.1 or later) from among DOS, PC-DOS or MS-DOS;
- a dot-matrix printer

If the computer is likely to be used for other applications, it is advisable to have 512 k or 640 k RAM, and 20 Mbytes or more of hard disk storage. A back-up streamer tape-drive should also be considered for data security in case of hard disk failure.

In countries where the electricity supply is unstable and protection of the electricity distribution system against lightning is inadequate, a surge protection device should be fitted to the power supply of the computer. Dehumidifiers and air-conditioning may be required to cope with extremes of humidity and temperature.

The cost of basic computing equipment varies from country to country, but is usually less than US \$6000. Operational expenses will also vary with the work-load, but the following items need to be considered:

Consumables

Paper for the printer

Ribbons (or ink etc.) for the printer

Diskettes

Electricity

Tapes (for streamer back-up: optional)

Optional hardware

Tape streamer device

Surge protection device

Other

Maintenance contract for the hardware

Installation, documentation and training

The installation of CANREG requires precise definition of the registry's requirements. These requirements are incorporated in a customized version of CANREG, which is particular to the registry concerned, yet compatible with other CANREG systems. One design criterion for CANREG was that it should be possible for the system to be modified, as required, by users who might have no programming skills at all. This requires an initial investment of a few days' work in which the data items to be collected are specified, the data entry screen and output styles are designed, the various dictionaries constructed or modified, and the whole system tested with real data. Each element of the system can subsequently be modified by the user.

A complete user guide is provided, and the data entry clerks are trained on site. A support service is then provided to deal with problems in the early period of operation, and to advise on other aspects of cancer registration, including analysis and presentation of the data.

Installed user base

By early 1990, CANREG had been installed and is in use in more than twenty cancer registries, mainly in developing countries of Africa, Asia, Latin America and Oceania.

Limitations and proposed developments in CANREG

The CANREG system has been especially designed for use on microcomputers, and cannot be used on mainframe computers. It is not suitable for use with data sets which are very large, or which contain more than 45 data items per record.

The system is under continuous review, and the various modifications either in progress or under consideration include:

- additional verification of data at entry of individual records, e.g., warnings or rejection for unlikely or impossible combinations of site and morphology;
- an option to carry out the same validity checks used when entering individual cases on data files about to be imported.
- an option to define a group of variables (e.g., name, date of birth, and tumour site) for use in searching for possible duplicate registrations in the whole data set;
- simplified, user-driven (interactive) installation procedures;
- selective back-up facilities for data sets (complete data, or only records added or modified since the last back-up), dictionaries and indexes.
- simplified procedures for designing or modifying the output styles;
- more extensive analysis and tabulation facilities, such as simple graphics of age-incidence curves, using only dot-matrix printers.

Obtaining the CANREG system

The system is made available primarily to member registries of the International Association of Cancer Registries, or to registries which collaborate with the International Agency for Research on Cancer. For registries in developing countries, assistance is usually provided to modify the program on site (see above), to ensure that codes, dictionaries and other elements of the system meet the user's requirements; then to install the system, and finally to train the staff who will use it.

Enquiries should be addressed to the Chief, Unit of Descriptive Epidemiology, International Agency for Research on Cancer, 150 cours Albert Thomas, 69372 Lyon Cédex 08, France.