

6. INPUT PROCEDURES

Notifications are received on registration/notification forms, or as computer files. Input procedures concern entering the information onto CanReg version 4 or 5 (Fig 6. 1).



Fig 6. 1 CanReg Software

The CanReg system allows input, storage, checking, back up and analysing cancer registry data. The input process also includes a number of inbuilt checks, to make sure that very obvious mistakes are flagged for correction. Incomplete or incorrect registrations cannot be CONFIRMED, and will remain in a pending state until corrected/completed.

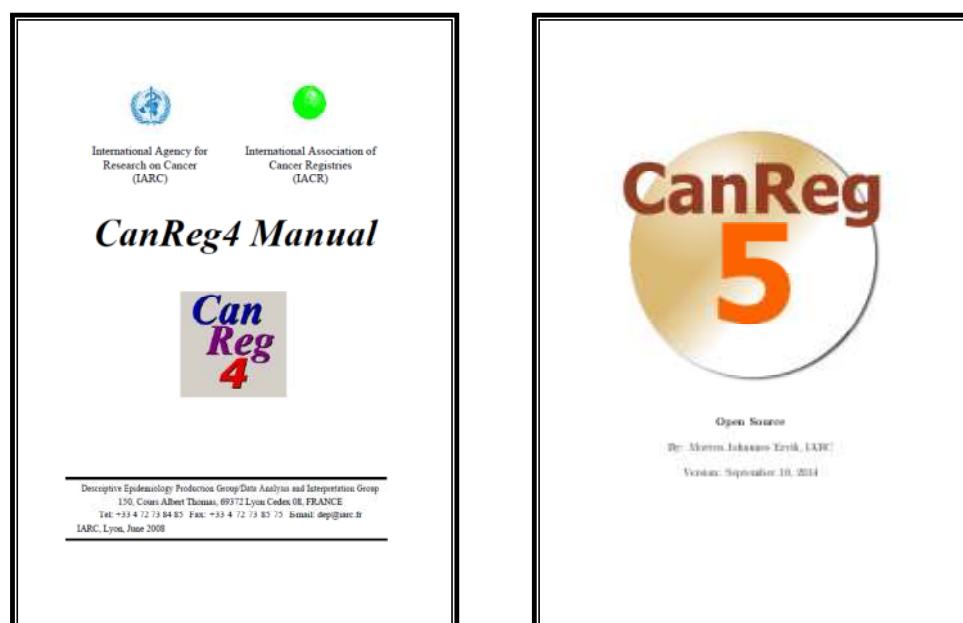


Fig 6. 2 CanReg User Guides

CanReg 4 stores one record for each cancer (tumour), noting for each cancer, all the separate SOURCES of notification; a special key (Multiple Primary Code) allows one to bring together different cancers for the SAME person.

CanReg 5 splits this information in three tables: Patient, Tumour and Source. For each patient, you can store as many tumour records as you need, and for each tumour you can store as many source records as you need.

The CanReg manuals (Fig 6. 2) give detailed instructions on data entry procedure, including checking to see if a given patient has already a record, allowing for updating existing records, and creating new ones. Fig 6. 3 shows the basic processes involved.

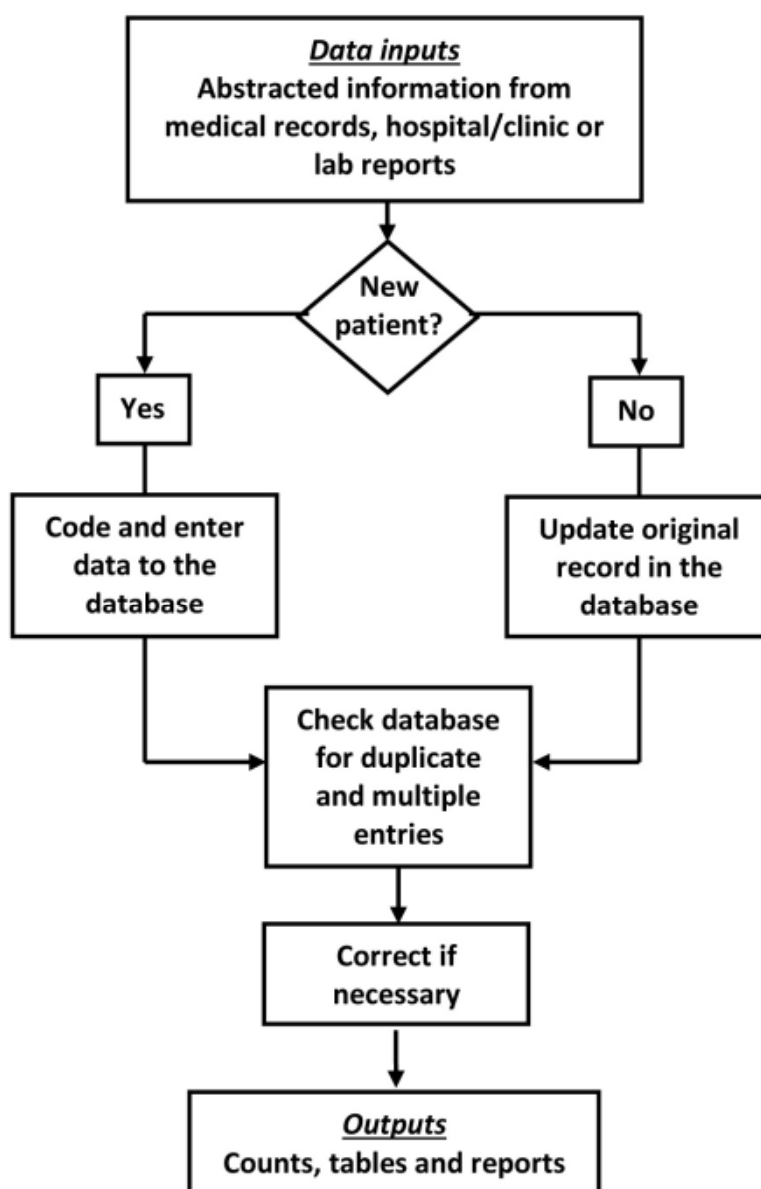


Fig 6. 3 Flow chart of data entry to database

DEATH CERTIFICATE NOTIFICATIONS

If the registration/notification form has been completed from a death certificate, leave **4 weeks** before data entry, to allow time to find the cases from any hospital sources.

Registration/notification forms completed from a death certificate (see section 4. ABSTRACTING) should be checked to see if the cancer case, or person, has already been registered, using the Browse / Edit function in CanReg in the same way as with a hospital abstract.

- ☞ If the case has been registered previously the record is updated with the date of death and any other new information.
- ☞ If there is no registration for the case; the place of death is checked. If the patient had died in hospital, the case should be “followed back” to see if the hospital record can be traced.
 - If it can be found - AND THE PATIENT REALLY DID HAVE CANCER - a registration/notification form should be completed from the hospital record with all the mandatory variables.
The case is registered with TWO sources (hospital and death certificate)
 - If there has been no previous registration and it proves impossible to trace any record of the case having been seen in hospital:

EITHER

- The case is registered as a new cancer using the information on the death certificate.
- Enter basis of diagnosis = 0 (Death Certificate Only)
- Set date of incidence = date of death (UNLESS there is information on date of diagnosis on the certificate).
- Source of information will be death certificate.

OR

If there is doubt about the accuracy of the cause of death statement (for example, the certificate has been issued by a non-medical person), the case not registered (e.g. left pending).

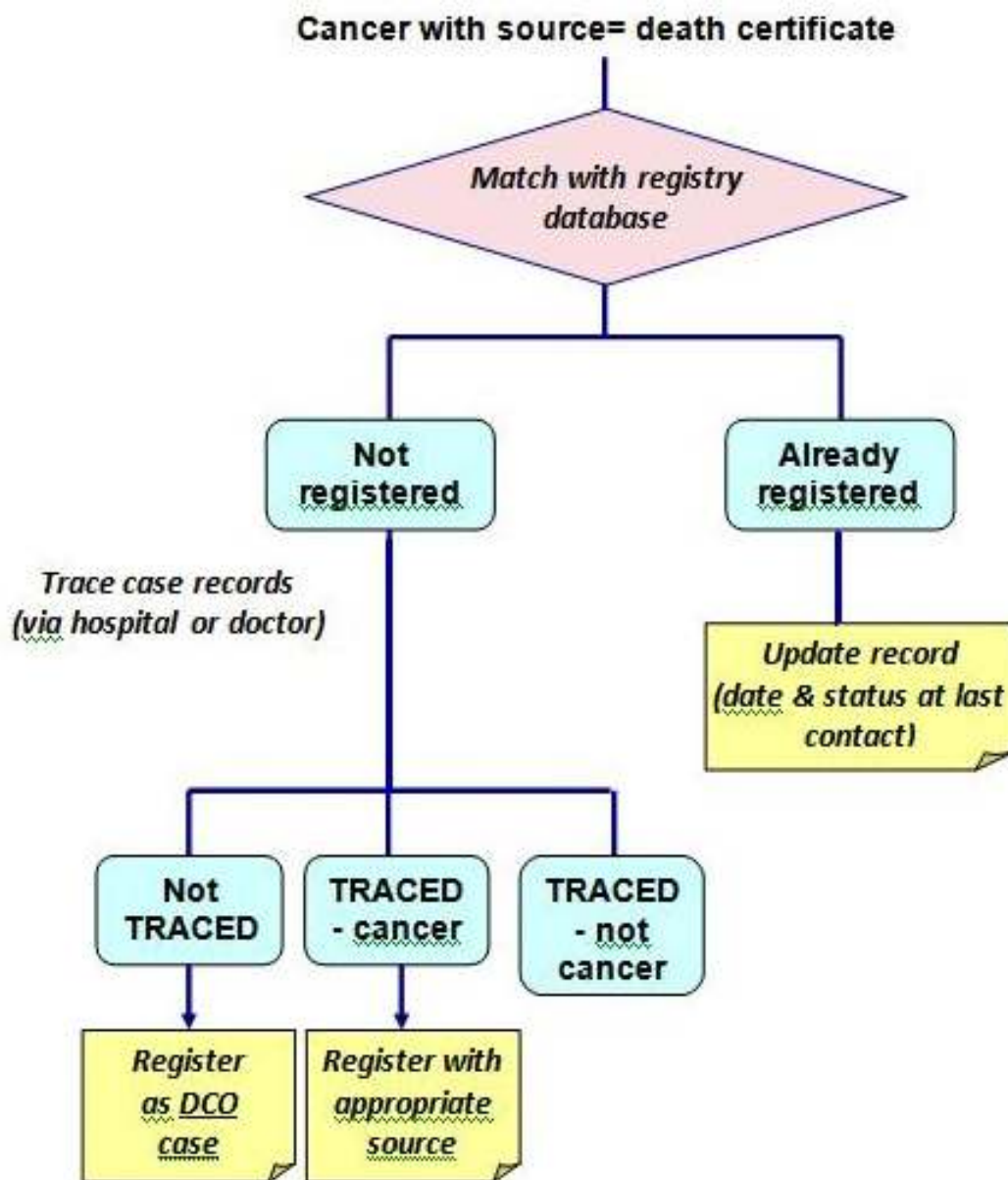


Fig 6. 4 The use of death certificates to identify new cases of cancer

7. DATA STORAGE

- Registration/notification forms must be filed numerically by registration number (Fig 7.1)
- Store in a cabinet that can be locked.
- The documents should be secure and inaccessible to unauthorized persons.
- They should be protected against loss or damage from fire, floods or any other interference.
- A BACK UP should be made of the CanReg database at the end of each day. The backup may be stored on a pen/flash drive/external portable hard drive/CD or other electronic media. This should be stored in a secure, locked cupboard or drawer that is secure and inaccessible to unauthorized persons.
- On transit data should be in a lockable suitcase.



Fig 7. 1 Filing of registration/notification forms

8. CONFIDENTIALITY

The Registry aims to maintain the confidentiality of all cancer information collected for the following reasons:

- To protect the privacy of the cancer patients
- To protect the privacy of the healthcare facilities reporting the cancer case
- To protect the privacy of the cancer patient's healthcare providers
- To protect from abuse and misuse of the cancer data

Guidelines on Confidentiality for Population-Based Cancer Registration have been published by IACR/IARC (IARC Internal Report No. 2004/03). They may be obtained on request from the AFCRN Secretariat.

Definition of confidential data

Confidential data include data that identifies specific information on the patient, healthcare facilities and healthcare providers reporting the case. The cancer registry should maintain the same standards of confidentiality as applicable to the confidentiality of medical records and clinician-patient relationship.

All staff must sign confidentiality document to preserve the anonymity of the registry data and not to divulge any information even after employment ceases. An example is provided in Appendix 5.

8.1 LOGISTICAL ASPECT OF CONFIDENTIALITY

The following measures to ensure confidentiality should be implemented:

Data collection

For data collected on registration/notification forms, it is the responsibility of the registry staff to preserve their confidentiality. Forms should be kept under lock and key preferably in a filing cabinet. They should not be left in a place where an unauthorized person might have access e.g. in your car.

Data transmission

When sending information through the mail:

- Use registered mail.
- Information should be sent in two separate lists; one of names and the other medical information which when in the registry are merged.
- Use double envelopes; the exterior one with a general address and the interior with the address to the authorized recipient who should ideally be the registry director or delegated/authorised person.

Confidential data should **NEVER** be sent by fax.

When information is sent electronically such as USB hard drives or CDs it is important to take measures to ensure that these will not get lost, and not be easily read by other parties. The following precautions may be taken:

- Encrypting of the names at various level.
- Preparation of a separate CDs or USB with the names and one with the tumour related data.
- Keep a record of all electronically transmitted and received data.
- Data not to leave the registry premises without authorization.

Computer

With data kept in computer user names and passwords should be used and changed regularly and it should be known only to the authorised users.

Telephone

Confidential information should **NEVER** be given over the telephone, nor should enquiries from collaborators concerning confidential data be given over the telephone.

8.2 ACCESS TO AND STORAGE OF DATA

Strict security measures should be exercised to ensure confidentiality. These include:

- Access to the registry should be limited and restricted to authorized persons only (Fig 8.1).
- All registry records should be stored in a room which can be locked and access limited only to authorized persons.
- Provide lockable filing cabinets
- Use shredder machine to destroy unwanted forms



Fig 8. 1 Restricted Entry Notice

8.3 USE AND RELEASE OF DATA

Confidential data may be provided by the registry only upon written request, (see Appendix 6) which should include the exact purpose for which the data will be used, the information required the name(s) of the person(s) responsible for keeping the confidential information and the time period for which the data are needed.

- The registry should make sure that those receiving the data:
 - Are bound by the same rules of confidentiality observed by the registry staff.
 - Will use the data only for the purpose agreed upon at the time of provision, and will not make them accessible to other parties.
 - Will destroy the data when they are no longer needed for the said purposes.
- No information should be provided to insurance companies, medical funds pension schemes employers, the police or to a physician having to examine an individual for such purpose.

Aggregate data

These kinds of data do not need strict confidentiality measures and include prepared tables, graphs and reports.

Individual data

Cancer registries contribute to investigations on the cause of cancer and the registry may frequently be asked to provide the names of patients with given cancers so that they can be included in a study. Patients' names may be disclosed to the treating physician. Otherwise, patients' names can be disclosed to the researchers who have the authority/approval of the registry director and the ethics committee. Names may be disclosed to researchers with the agreement that the patients or members of the family may not be identified or any detailed information which permits any form of identification.

International release

When sending data abroad the registry staff should ensure that patients' identifications are not disclosed. Cases may be identified by a code number or patients' registry number (which can be linked to the registration record within the registry).

Requests by researchers for data from AFCRN members in more than one country should be referred to the AFCRN Research Committee.

8.4 DISSEMINATION OF DATA TO ORGANISATIONS OUTSIDE RESEARCH for example THE PRESS

Only the registry director may release of data to the media. He/she should insist on viewing the draft of the article prior to release or publication. Identifiable data should **NEVER** be released to the media.

9. QUALITY CONTROL

The primary goal of a population-based cancer registry is to determine the incidence of cancer within its geographical population. It is therefore of the utmost importance that the registry data be of good quality. This means that the information gathered, especially on essential items should be complete, consistent and accurate, and that coverage of the population should be as complete as possible. Quality control concerns three aspects of registry work:

VALIDITY: This is the accuracy of the information registered (or, the proportion of cases recorded as having a given characteristic that truly have that attribute).

COMPLETENESS: This is the extent to which all of the new (incident) cancers occurring in the target population of the registry are included in the database.

TIMELINESS: the speed with which registry data is ready for analysis and reporting.

9.1 MEASURING VALIDITY (ACCURACY) OF REGISTRY DATA

The methods used are as follows:

1. Re-abstracting and recoding “audits”
2. Reporting “Morphology Verified” percentages
3. Reporting DCO percentages
4. Reporting on percentage of missing information
5. Internal consistency checks

9.1.1 Re-abstracting and recoding audits

Re-abstracting audits and recoding audits often are used to assess the accuracy (agreement with source medical records) and reproducibility (agreement among data collectors) of registry data.

They need to be performed by an auditor – either from the registry (for example, the Director, or Registry Manager), or an “expert” consultant from outside.

The objective of a re-abstracting study is to measure the level of agreement between data in the registry and data re-abstracted and recoded by the auditor from source records (the hospital medical records for most cases).

Re-abstracting

A sample of registrations is selected from the registry database by the auditor. Eligible cases are those diagnosed at least one year prior to the year of the study.

He/she will select the sample:

- at random from the whole database
- randomly from certain sources that are known to cause problems to the registry staff
- randomly, but with the same number of cases drawn for each registrar

The sample will be for registrations from a single year (or period of a few years) that are subject of the quality control exercise.

Hilsenbeck et al, (1987), of the Centralized Cancer Patient Data System in the USA suggested that the sample size should be, as a minimum, 3-4 cases per registrar per month.

For these registrations, the records from which the case was abstracted are requested from the source concerned. This means sending a list of the case records required (the list contains case number, patient's name, date) to the sources (hospital records departments, for example) and requesting that the case files are ready for the exercise.

The auditor will then abstract the case onto the registration form (WITHOUT looking at the original registration). The re-abstracts are compared with the original (either the registration form, or the details from the CanReg database).

For each re-abstracted data item, the auditor's codes are compared to the original codes to identify discrepancies. If the codes do not match, the discrepancy is classified as to severity according to major and minor discrepancy definitions (see Table 9.1). The Table 9.2 shows an example of results of such a study.

Item	Code	Major disagreement	Minor disagreement
<u>Demographic</u>			
Sex		any difference	
Age		>1 years difference	difference \geq 3 months
Birthdate	dd/mm/yyyy	different yyyy	difference in month/day
Ethnic group			any difference
Place of residence		in/out of registry area	any
<u>Tumour</u>			
Date of incidence	dd/mm/yyyy	different yyyy	difference \geq 3 months
Primary site	ICD-O (Cxx.y)	difference in xx	difference in y (3rd digit)
Morphology	ICD-O (Mxxxy)	difference in xxx	difference in y (4th digit)
Behaviour	ICD-O	any difference	
Basis of diagnosis		difference MV or non-MV or DCO	difference within MV difference within non-MV
Laterality			any difference
Stage		difference resulting in change of UICC stage (I-IV)	any other difference
<u>Treatment</u>			
Type: surgery radiotherapy chemotherapy hormone therapy		given v not given	any different code (including 9 [unknown])
Date		difference \geq 1 month	difference < 1 month
<u>Follow up</u>			
Date of last contact	dd/mm/yyyy	difference \geq 3 months	difference < 3 months
Status at last contact		any difference	

Table 9. 1 Major and minor disagreements for selected key data items

Data Items	Data Items Reabstracted	Number in agreement	% agreement
Sex	50	50	100%
Race	50	48	96%
Age	50	47	94%
Date of Diagnosis	50	43	86%
Primary Site	50	46	92%
Histology	50	46	92%
Basis of diagnosis	50	48	96%
Stage	50	33	66%
Treatment			
Surgery	50	48	96%
Radiation Therapy	50	47	94%
Chemo-Endocrine Therapy	50	46	92%
Other Therapy	50	50	100%
Date of Treatment	50	45	90%
Date of Last Contact	50	48	96%
Vital Status at Last Contact	50	49	98%
TOTALS	750	694	93%

Table 9. 2 Results of a Hypothetical Re-abstracting Study.

Recoding audits These look at the level of agreement between registry staff and the auditor for records already in the registry. The auditor uses the text contained on the registration form to recode a sample of actual case records in the registry database.

As in a re-abstracting study, for each recoded case, codes for each data item are compared for discrepancies with those assigned by the auditor. These studies show:

- The types of tumour records in which discrepancies occur more frequently.
- Sources of variation (e.g., misinterpretation of source document information, information not available at initial abstracting, misinterpretation of coding rules, inadequate or erroneous consolidation of data between records).
- Effect of misclassification that could affect data analysis and use (e.g., are tumours more frequently over-staged or under-staged?).
- Data quality with respect to other factors such as who collects the data (permanent registrars versus medical staff), training and skills of the registrars collecting the data, and difficulty of abstracting and coding the specific data items.

This information should be used to identify training needs and to modify registry processes and procedures to ensure future improvement in data quality.

9.1.2 Percentage of cases with a morphologically verified diagnosis (MV%)

Morphological verification refers to cases for which the diagnosis is based on histology or cytology.

Procedure:

For the time period for which the quality control exercise is being performed (for example, one year, three years, 5 years), make a table, with, for each sex, the number of cases , by cancer site (using the ICD-10 codes) for each “Basis of Diagnosis” code (see Table 9.3, left side).

Then, group together the “basis of diagnosis codes” that represents diagnoses based on examination by microscope (generally in pathology or haematology labs). The codes (section 5.4, page 26) are:

5. Cytology or haematology
6. Histology of a metastasis
7. Histology of a primary tumour

The MV% is the percentage of all registrations with these “basis” codes.

The right hand side of Table 9.3 shows how the codes (ICD-10) for cancer site, and for “basis of diagnosis” can be grouped (with Basis of diagnosis as DCO/ Clinical/ M.V.) in a table suitable for publication in a registry report.

Basic data: Table of site (ICD-10) by basis of diagnosis													
SITE (ICD-10)	BASIS CODES										Total	BASIS	
	0	1	2	3	4	5	6	7	8	1-4		5-8	
0	0	3	0	0	0	0	0	0	0	0	3	3	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	4	0	0	0	0	0	0	11	0	15	4	11
3	0	1	0	1	0	0	0	0	5	0	7	2	5
4	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	4	0	0	0	0	0	0	6	0	10	4	6
6	1	5	0	1	0	0	0	0	3	0	10	6	3
7	0	5	0	0	0	0	0	0	6	0	11	5	6
8	0	1	0	0	0	0	0	0	3	0	4	1	3
9	0	2	0	0	0	0	0	0	5	0	7	2	5
10	0	4	0	0	0	0	0	0	6	0	10	4	6
11	0	22	0	0	0	0	0	1	27	0	50	22	28
12	0	0	0	0	0	0	0	0	1	0	1	0	1
13	0	2	0	0	0	0	0	0	1	0	3	2	1
14	0	2	0	0	0	0	0	0	0	0	2	2	0
15	16	95	0	0	0	0	0	0	71	0	182	95	71
16	8	33	0	1	0	0	0	2	37	0	81	34	39
17	0	0	0	0	0	0	0	0	3	0	3	0	3
18	1	27	0	0	0	0	0	0	22	0	50	27	22
19	0	7	0	0	0	0	0	0	5	0	12	7	5
20	4	24	0	0	0	0	0	0	34	0	62	24	34
21	0	2	0	0	0	0	0	0	4	0	6	2	4
22	8	104	1	0	0	0	0	0	87	0	200	105	87
23	0	0	0	0	0	0	0	0	4	0	4	0	4
25	3	27	0	0	0	0	0	0	6	0	36	27	6
26	0	0	0	0	0	0	0	0	2	0	2	0	2
30	0	8	0	0	0	0	0	0	11	0	19	8	11
31	0	1	0	0	0	0	0	0	2	0	3	1	2
32	0	13	0	0	0	0	0	0	10	0	23	13	10
34	2	37	0	0	0	0	0	0	20	0	59	37	20
35	0	0	0	0	0	0	0	0	0	0	0	0	0

Cancer site	ICD-10	Basis of diagnosis				
		No. Cases	(% total)	DCO	Clinical	M.V.
Oral cavity & pharynx	C00-C14	133	3.1	0.8%	42.9%	56.4%
Oesophagus	C15	182	4.3	8.8%	52.2%	39.0%
Stomach	C16	81	1.9	9.9%	42.0%	48.1%
Large bowel	C18-C21	130	3.1	3.8%	46.2%	50.0%
Liver	C22	200	4.7	4.0%	52.5%	43.5%
Pancreas	C25	36	0.9	8.3%	75.0%	16.7%
Larynx	C32	23	0.5	0.0%	56.5%	43.5%
Lung	C33-C34	59	1.4	3.4%	62.7%	33.9%
Bone	C40-C41	54	1.3	0.0%	42.6%	57.4%
Melanoma of Skin	C43	25	0.6	0.0%	28.0%	72.0%
Other Skin	C44	54	1.3	0.0%	37.0%	63.0%
Kaposi sarcoma	C46	1035	24.4	1.1%	25.2%	73.7%
Breast	C50	334	7.9	0.9%	53.3%	45.8%
Cervix Uteri	C53	402	11.6	2.6%	43.5%	53.9%
Corpus Uteri	C54	33	0.8	3.0%	27.3%	60.7%
Ovary	C56	57	1.3	5.3%	40.1%	45.6%
Prostate	C61	236	5.6	1.3%	33.9%	64.8%
Kidney	C64	55	1.3	0.0%	38.2%	61.8%
Bladder	C67	22	0.5	0.0%	50.0%	50.0%
Eye	C69	122	2.9	0.0%	15.6%	84.4%
Brain, Nervous system	C70-C72	36	0.9	8.3%	61.1%	30.6%
Thyroid	C73	35	0.8	2.9%	14.3%	82.9%
Hodgkin disease	C81	46	1.1	0.0%	15.2%	84.8%
Non-Hodgkin lymphom	C82-C85,C	298	7.0	1.3%	45.0%	53.7%
Leukaemia	C91-C95	98	2.3	6.1%	40.8%	53.1%
All sites Total	All	4235	100	2.2%	39.2%	58.6%

Table 9. 3 Example of calculation of MV% (Registry X, data for 2005-2007)

One of the standard tables in CanReg5 (“Data Quality Indicators”) includes the MV% - in addition to other indicators of data quality (see Table 9.4).

Training System (English) (2001–2005)							
Data Quality Indicators							
MALE							
SITE	Cases	% Total	ASR(se)	MV(%)	CLIN(%)	DCO(%)	ICD10
Mouth & pharynx	418	5.47	16.62 (0.84)	98.80	0.72	0.48	C00–14
Oesophagus	197	2.58	8.31 (0.61)	88.83	2.03	9.14	C15
Stomach	430	5.63	19.11 (0.95)	91.63	2.09	6.28	C16
Colon, rectum, anus	529	6.93	22.45 (1.01)	93.38	1.32	5.29	C18–21
Liver	116	1.52	5.09 (0.49)	60.34	8.62	31.03	C22
Pancreas	81	1.06	3.60 (0.41)	60.49	11.11	28.40	C25
Larynx	163	2.13	7.11 (0.57)	95.09	3.68	1.23	C32
Lung, trachea, bronchus	500	6.55	23.14 (1.05)	79.80	6.60	13.60	C33–34
Pleura & other thoracic	20	0.26	0.67 (0.16)	70.00	15.00	15.00	C37–38
Melanoma of skin	122	1.60	4.84 (0.46)	96.72	0.82	2.46	C43
Prostate	2153	28.19	105.44 (2.29)	95.45	1.49	3.07	C61
Testis	56	0.73	1.41 (0.20)	92.86	3.57	3.57	C62
Kidney & urinary NOS	132	1.73	5.52 (0.50)	91.67	2.27	6.06	C64–66,68
Bladder	265	3.47	12.42 (0.77)	94.34	2.26	3.40	C67
Brain & nervous system	211	2.76	7.16 (0.53)	82.46	5.21	12.32	C70–72
Thyroid	65	0.85	2.08 (0.28)	98.46	0.00	1.54	C73
Ill-defined	204	2.67	8.92 (0.64)	63.24	12.25	24.51	C76–80
Lymphoma	414	5.42	15.93 (0.82)	87.44	2.17	10.39	C81–85,90,88,96
Leukaemia	226	2.96	7.91 (0.56)	76.55	1.33	22.12	C91–95
All sites but C44	6727	88.08	293.72 (3.70)	89.58	2.76	7.66	ALLbC44
FEMALE							
SITE	Cases	% Total	ASR(se)	MV(%)	CLIN(%)	DCO(%)	ICD10
Mouth & pharynx	139	1.79	4.45 (0.39)	94.24	3.60	2.16	C00–14
Oesophagus	61	0.78	2.20 (0.29)	90.16	1.64	8.20	C15
Stomach	278	3.58	9.40 (0.58)	92.81	2.16	5.04	C16
Colon, rectum, anus	608	7.82	20.70 (0.86)	93.91	1.48	4.61	C18–21
Liver	53	0.68	1.80 (0.25)	47.17	3.77	49.06	C22
Pancreas	104	1.34	3.68 (0.37)	57.69	9.62	32.69	C25
Larynx	36	0.46	1.28 (0.22)	97.22	2.78	0.00	C32
Lung, trachea, bronchus	308	3.96	11.11 (0.64)	78.25	5.84	15.91	C33–34
Pleura & other thoracic	11	0.14	0.31 (0.10)	81.82	0.00	18.18	C37–38
Melanoma of skin	124	1.60	3.80 (0.36)	100.00	0.00	0.00	C43
Breast	1766	22.72	54.78 (1.35)	97.06	1.30	1.64	C50
Cervix	904	11.63	26.70 (0.93)	98.23	0.55	1.22	C53
Corpus & Uterus NOS	226	2.91	7.78 (0.53)	98.23	0.88	0.88	C54–55
Ovary & adnexa	228	2.93	7.24 (0.50)	92.11	2.19	5.70	C56
Kidney & urinary NOS	93	1.20	3.30 (0.35)	92.47	2.15	5.38	C64–66,68
Bladder	121	1.56	4.29 (0.40)	91.74	1.65	6.61	C67
Brain & nervous system	159	2.05	4.61 (0.38)	78.62	4.40	16.98	C70–72
Thyroid	366	4.71	9.99 (0.55)	96.99	1.64	1.37	C73
Ill-defined	173	2.23	5.88 (0.46)	70.52	10.40	19.08	C76–80
Lymphoma	380	4.89	11.98 (0.64)	89.21	2.11	8.68	C81–85,90,88,96
Leukaemia	234	3.01	6.93 (0.47)	84.62	0.43	14.96	C91–95
All sites but C44	6750	86.85	214.37 (2.70)	92.00	2.09	5.91	ALLbC44

Cases of unknown age (21 M / 26 F) were excluded from these analyses

Table 9. 4 Output of CanReg-5 (Data Quality Indicators)

This MV% is traditionally considered as a sort of “gold standard”, with suspicion falling upon the accuracy of diagnosis by other means (although in reality a diagnosis based on an MRI or CT scan may be just as accurate as one based on exfoliative cytology). A high MV% is taken to mean accuracy of diagnosis, whereas a low MV% casts doubt on the validity of the data.

The absolute value of the MV% needs to be compared with an “expected” value that is reasonable given the circumstances (state of medical technology, local clinical practice) in which the registry operates. Therefore, the MV values (by site and, preferably also by sex) should be compared with an appropriate set of standards, so that values that are significantly different can be identified.

Table 9.5 provides the “standard” values of MV% for sub-Saharan Africa, with which your own values can be compared⁵.

ICD-10 code	Cancer site	Male	Female
		MV%	MV%
C00–14	Oral cavity and pharynx	68.6	71.4
C15	Oesophagus	46.7	45.9
C16	Stomach	53.1	53.4
C18–21	Large bowel	62.1	61.3
C22	Liver	11.7	12.6
C25	Pancreas	16.8	22.2
C32	Larynx	66.2	73.3
C33–34	Trachea, bronchus, and lung	44.7	64.1
C43	Melanoma of skin	76.9	90.0
C50	Breast	66.7	66.1
C53	Cervix uteri	0.0	62.4
C54–55	Corpus uteri, uterus unspecified	0.0	64.6
C56	Ovary	0.0	51.3
C61	Prostate	59.8	0.0
C62	Testis	48.3	0.0
C64–66	Kidney, renal pelvis, and ureter	68.8	67.1
C67	Bladder	39.7	45.0
C70–72	Brain, central nervous system	51.5	41.8
C73	Thyroid	65.4	73.8
C81–88, C90	Lymphomas	84.5	82.0
C91–95	Leukaemia	87.2	88.4
C76–80	Unspecified	48.4	39.8
C00–96 (excluding C44)	All sites (excluding non-melanoma skin)	57.4	61.1

MV%, percentage of cases with a morphologically verified diagnosis.
^a The Gambia (1997–1998), Mali, Bamako (1994–1996), Uganda, Kyadondo County (1993–1997), Zimbabwe, Harare: African (1993–1997).

Table 9. 5 Mean values of MV% for cancer registries in sub-Saharan Africa

The CanReg5 Table (“Data Quality Indicators”) does not yet show whether the recorded MV% is significantly different from (higher or lower) this standard (see Table 9.4).

Whereas a MV% significantly lower than the expected value may give rise to concern about a lack of validity, it is generally not the cancer registry that can influence the availability of, or use of, pathology services within its area. Usually, in Africa, the opposite situation – a relatively high MV% – is cause for concern. This is because collecting data on cancer cases from pathology departments is much easier than trawling through clinical services or ill-organized hospital archives. A large proportion of cases diagnosed via the pathology

⁵ A suitable statistical test is has been described in Bray & Parkin (2009)

department may well suggest defects in case finding and, hence, incomplete registration. Worse, the incompleteness will be biased, with the database containing a deficit of cancers that are not easy to biopsy, and so are diagnosed by other methods (e.g. lung, liver, brain, and pancreatic cancer).

9.1.3 Percentage of cases for which the only information came from a death certificate (DCO%)

DCO cases are those registered on the basis of information on a death certificate, and for which no other information could be traced. As described earlier (section 6.1), the nature of death certificates in Africa varies widely, from those issued as part of a civil registration of vital events to those generated in a hospital mortuary.

However, almost always the accuracy of the diagnostic information is questionable, since the person writing out the certificate may have had little contact with the patient before death and may be ill-informed about how to record cause of death. They may even have no medical training at all. Thus, if no other clinical record for persons who apparently died of (or with) cancer can be found, there is a reasonable suspicion that the diagnosis was simply wrong.

If you include such cases in the database, and if they comprise a large proportion of cases, the validity of the data is suspect.

Procedure:

As for MV% (see 9.1.2), for the time period for which the quality control exercise is being performed (for example, one year, three years, 5 years), make a table, with, for each sex, the number of cases, by cancer site (using the ICD-10 codes) for each “Basis of Diagnosis” code.

The DCO cases are those with *basis of diagnosis* = 0

See Table 9.3

The DCO% is the percentage of all registrations with this “basis” code (=0)

As for MV%, we calculate DCO% by cancer site, and, ideally, by sex.

The CanReg5 Table (Data Quality Indicators) shows the percentage of DCO cases, by site and sex (see Table 9.4).

What is an Acceptable Level of DCO% ?

This is difficult – it depends on local circumstances, for example availability of death certificates, success in record linkage, accuracy of cause of death statements on the certificate.

Some collections of cancer registry results have proposed more or less arbitrary standards; for example, Cancer Incidence in Five Continents Volume IX (Curado et al, 2007) considered <10% DCO to be category “A” for quality, and 10-20% category “B”. The criteria in the North American Association of Central Cancer Registries (NAACCR) is a DCO of less than 3 percent for “gold” standard and less than 5 percent for silver (Hofferkamp, 2008).

9.1.4 Proportion (or percentage) of cases with missing data

The proportion of cases with unknown values of different data items, such as age or stage, is also an indicator of data quality. The data items that should be assessed for missing values are:

- ☞ Age
- ☞ Primary site
- ☞ Stage
- ☞ Follow up

Note that it is **NEVER** acceptable for “sex” to be missing.

Unknown values can result from problems with the registration process, may also result from inadequate case histories or investigation, or ambiguity in the medical record.

For “AGE” we wish to calculate the number of registrations (by cancer site and sex) for which age was recorded as unknown ([code 99](#)).

Primary site uncertain (PSU%) includes, in addition to “Unknown Primary Site” (C80 in ICD-10), other rubrics (e.g. malignant neoplasms of ill-defined organs of the digestive system (ICD-10 C26), respiratory system (C39) endocrine system (C75), and peritoneal and retroperitoneal neoplasms (C48) as well as those of “Other and Ill-defined Sites” (C76).

The standard Tables produced by CanReg show the numbers of cases with Age Unknown, and with a row entitled “Other and Unspecified” (O & U) (Fig 9.6).

A high proportion of cases assigned to the O&U/PSU category means there is low accuracy of diagnosis, usually due to the failure to specify the site of the primary cancer in cases diagnosed on the basis of tissue obtained from a metastasis. Incidence rates for cancers at specific sites will be underestimated if a large proportion of registered cases appear in the “Other and/or Unspecified” category, rather than with their true diagnosis.

As for DCO%, some collections of cancer registry results have proposed more or less arbitrary standards for % missing; for example, CI5 Volume IX proposed acceptable maxima for the percentage of cases with age unknown (<20%), and ill-defined sites (<20%) (Curado et al, 2007). The NAACCR standards include <3% cases with Age missing and <5% cases with unknown primary site (Hofferkamp, 2008).

Gambia National Cancer Registry (2007-2009)
Gambia National Cancer Registry 2007-2009
Cases by age group (Period) - Male

SITE	ALL AGES	AGE UNK	0	5	10	15	20	25	30	35	40	45	50	55	60	65+	(%)	ICD- (1986)
Lip	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0	C80
Tongue	2	1	-	-	-	-	-	-	-	-	1	-	-	-	-	-	0.3	C81-82
Mouth	5	4	-	-	-	-	-	-	-	-	-	-	-	1	-	-	0.6	C83-86
Salivary glands	1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0.1	C87-88
Tonsil	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0	C89
Other oropharynx	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0	C90
Nasopharynx	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0	C91
Hypopharynx	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0	C92-93
Pharynx unspecified	1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.1	C94
Oesophagus	13	2	-	-	-	-	1	-	3	1	-	1	2	1	-	2	1.6	C95
Stomach	22	4	-	-	-	-	-	-	4	1	1	1	4	1	3	4	2.8	C96
Small intestine	3	0	-	-	-	-	-	-	1	-	-	-	-	2	-	-	0.4	C97
Colon	1	0	-	-	-	-	-	-	-	-	-	-	1	-	-	-	0.1	C98
Rectum	14	1	-	-	-	2	-	-	1	1	2	1	-	3	1	2	1.8	C99-100
Anus	1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.1	C101
Liver	491	40	3	1	2	18	20	52	54	51	42	51	29	41	29	66	81.7	C22
Gallbladder etc.	1	0	-	-	-	-	-	-	-	-	-	-	1	-	-	-	0.1	C23-24
Pancreas	89	0	-	-	-	-	-	1	-	-	1	1	3	-	2	2	1.3	C25
Nose, sinuses etc.	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0	C30-31
Larynx	1	0	-	-	-	-	-	-	-	-	-	-	-	-	1	-	0.1	C32
Trachea, bronchus and lung	28	1	-	-	-	-	-	-	2	-	-	3	1	3	1	11	3.9	C33-34
Other thoracic organs	4	2	-	-	-	-	1	-	-	-	-	-	-	-	-	1	0.3	C37-38
Breast	18	7	-	-	-	-	2	1	2	-	1	1	2	-	1	1	2.3	C40-41
Melanoma of skin	4	2	-	1	-	-	-	-	-	-	-	-	-	-	-	1	0.5	C42
Other skin	2	4	-	-	-	2	-	-	-	2	-	-	-	-	2	2	1.0	C43
Mesothelioma	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0	C45
Kaposi sarcoma	6	2	-	-	-	-	-	-	2	-	-	1	-	-	-	1	0.8	C46
Connective and soft tissue	8	2	-	1	1	-	-	-	-	-	-	-	1	1	-	2	1.0	C47, C49
Brain	1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0.1	C39
Brain	2	0	-	-	-	-	-	-	1	-	-	-	-	-	-	1	0.1	C80
Prostate	80	7	-	-	-	-	1	-	-	1	-	5	1	1	7	27	6.1	C51
Testis	1	0	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0.1	C52
Other male genital organs	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0	C53
Kidney	4	1	-	-	-	-	-	1	-	-	1	-	-	-	1	-	0.8	C54
Renal pelvis	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0	C55
Ureter	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0	C56
Bladder	22	1	-	-	-	-	-	-	2	-	-	1	2	-	3	3	2.3	C57
Other urinary organs	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0	C58
Eye	5	0	3	1	-	-	-	-	-	-	-	-	-	-	-	1	0.6	C60
Brain, nervous system	4	2	1	-	-	-	-	-	1	-	-	-	-	-	-	-	3.3	C70-72
Thyroid	4	1	-	-	-	-	-	1	-	-	-	-	1	-	-	1	0.5	C73
Adrenal gland	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0	C74
Other endocrine	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0	C75
Hodgkin disease	4	0	-	-	1	-	1	-	1	1	-	-	-	-	-	-	0.5	C81
Non-Hodgkin lymphoma	52	19	5	6	4	4	1	-	4	2	1	1	1	2	3	1	8.5	C82-83, C90
Immunoproliferative diseases	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0	C85
Multiple myeloma	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0	C86
Lymphoid leukaemia	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0	C91
Myeloid leukaemia	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0	C91-94
Leukaemia unspecified	5	0	-	-	2	-	-	-	-	-	1	1	-	-	-	1	0.6	C95
Myeloproliferative disorders	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0	M5D
Myelodysplastic syndrome	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0	M5E
Other and unspecified	18	4	1	-	-	2	-	2	1	1	-	1	5	1	-	-	2.1	C6-27
All sites	804	117	13	10	10	27	23	58	71	87	51	49	57	62	53	132		ALL
All sites but C44	796	113	13	10	10	26	27	58	71	86	51	49	57	62	52	131	100.0	ALL-C44

Table built on Jan 04 10:25:28 GMT 2015 by CanReg5

Table 9. 6 Standard Table (Cases by age group (period)) from CanReg5 showing the percentage of cases with age unknown (AGE UNK) by site, and % with "Other and Unspecified" sites (by age group)

9.1.5. Consistency checks

In computerized registries some aspects of validity of registered data are checked using automated routines. This is done when the data are being entered into CANREG, or as a part of a batch operation (off-line). A 'scale of errors' is set up in the system such that major errors result in complete rejection of a registration, while less serious ones are flagged to indicate that they contain an error. These cases must be "Confirmed" in CanReg, otherwise, they remain as "Pending" cases, and will not appear in analytic tables.

The most basic edit check is on the validity of the codes used, so that records with coded values outside the permitted range for the item (as defined in the registry) are rejected.

At the next level are checks of logical consistency between data items. A cancer cannot be diagnosed before the date of birth of a patient, a man cannot have ovarian cancer, and treatment cannot be undertaken for a patient who has died.

An edit program rejects these impossible combinations. It may also flag unlikely or unusual combinations such as those site-specific morphology terms which have only one possible topography code e.g. nephroblastoma which arises from the kidney should have a topography code C64.9 and hepatoma which arises from the liver should have a topography code of C22.0.

The IARC-CHECK program is incorporated into CanReg, and also available to check files of cases “in batch” – that is, outside of CanReg - is available from

http://www.iacr.com.fr/index.php?option=com_content&view=category&layout=blog&id=68&Itemid=445

It checks data for validity and consistency. The data items checked by the program are:

- registration number
- date of incidence
- age (or date of birth)
- sex
- site
- histology
- basis of diagnosis

The edit checks carried out by the program are described below:

1. Individual data item edits

Date of birth	Must be a valid date according to the format specified.
Incidence date	Must be a valid date according to the format specified.
Age at incidence	Must be a positive numeric value, not greater than 105 years. Age can be calculated if both birth and incidence dates are provided.
Sex	Must be a valid code
Site	Must be a valid ICD-O-3 code.
Morphology	Must be a valid ICD-O-3 code.
Behaviour	Must be 0, 1, 2, 3

2. Data combination edits

Incidence/birth dates	The birth date must be before the incidence date.																																				
Age/incidence/birth dates	When all are present, the calculated age must be equal to age +/- one year.																																				
Age/site/histology	<p>At certain ages some tumours are very unlikely. This routine identifies such sites and histologies:</p> <p>1. If the given age is less than 15, it performs a childhood check following rules described in the '<i>International Classification of Childhood Tumours</i>', page 11:</p> <table> <tr> <th>Diagnostic group</th><th>Unlikely age (years)</th></tr> <tr> <td>Hodgkin lymphoma</td><td>0-2</td></tr> <tr> <td>Neuroblastoma</td><td>10-14</td></tr> <tr> <td>Retinoblastoma</td><td>6-14</td></tr> <tr> <td>Wilms' tumour</td><td>9-14</td></tr> <tr> <td>Renal carcinoma</td><td>0-8</td></tr> <tr> <td>Hepatoblastoma</td><td>6-14</td></tr> <tr> <td>Hepatic carcinoma</td><td>0-8</td></tr> <tr> <td>Osteosarcoma</td><td>0-5</td></tr> <tr> <td>Chondrosarcoma</td><td>0-5</td></tr> <tr> <td>Ewing sarcoma</td><td>0-3</td></tr> <tr> <td>Non-gonadal germ cell</td><td>8-14</td></tr> <tr> <td>Gonadal carcinoma</td><td>0-14</td></tr> <tr> <td>Thyroid carcinoma</td><td>0-5</td></tr> <tr> <td>Nasopharyngeal carcinoma</td><td>0-5</td></tr> <tr> <td>Skin carcinoma</td><td>0-4</td></tr> <tr> <td>Carcinoma, NOS</td><td>0-4</td></tr> <tr> <td>Mesothelial neoplasms (M905_)</td><td>0-14</td></tr> </table> <p>2. If the given age is greater than 15, then the following combinations are considered to be unlikely:</p> <ul style="list-style-type: none"> • If age is less than 40 and site is C61._ and histology is 814_ • If age is less than 20 and site is: C15._,C19._,C20._,C21._,C23._,C24._,C38.4,C50._,C53._,C54._ or C55._ • If age is less than 20 and site is C17._ and histology less than 9590 (i.e. not lymphoma) <ul style="list-style-type: none"> • If age is less than 20 and site is C33._ or site is C34._ or site is C18._ and histology is not equal to 824_ (i.e. not carcinoid). • If age is greater than 45 and site is C58._ and histology is 9100 • If age is less than or equal to 25 and histology is 9732 or 9823 • If histology is 8910,8960,8970,8981,8991,9072,9470,951_ or 9687 	Diagnostic group	Unlikely age (years)	Hodgkin lymphoma	0-2	Neuroblastoma	10-14	Retinoblastoma	6-14	Wilms' tumour	9-14	Renal carcinoma	0-8	Hepatoblastoma	6-14	Hepatic carcinoma	0-8	Osteosarcoma	0-5	Chondrosarcoma	0-5	Ewing sarcoma	0-3	Non-gonadal germ cell	8-14	Gonadal carcinoma	0-14	Thyroid carcinoma	0-5	Nasopharyngeal carcinoma	0-5	Skin carcinoma	0-4	Carcinoma, NOS	0-4	Mesothelial neoplasms (M905_)	0-14
Diagnostic group	Unlikely age (years)																																				
Hodgkin lymphoma	0-2																																				
Neuroblastoma	10-14																																				
Retinoblastoma	6-14																																				
Wilms' tumour	9-14																																				
Renal carcinoma	0-8																																				
Hepatoblastoma	6-14																																				
Hepatic carcinoma	0-8																																				
Osteosarcoma	0-5																																				
Chondrosarcoma	0-5																																				
Ewing sarcoma	0-3																																				
Non-gonadal germ cell	8-14																																				
Gonadal carcinoma	0-14																																				
Thyroid carcinoma	0-5																																				
Nasopharyngeal carcinoma	0-5																																				
Skin carcinoma	0-4																																				
Carcinoma, NOS	0-4																																				
Mesothelial neoplasms (M905_)	0-14																																				
Site/histology	This routine identifies the morphological codes which are used exclusively with specific sites, or combinations of site and morphology which are unusual or unlikely. The morphological codes are grouped into 'families'.																																				
Sex/site	<p>Some sex/site combinations are impossible:</p> <ol style="list-style-type: none"> 1. If sex is male and site is: C51._,C52._,C53._,C54._,C55._,C56._,C57._ or C58._ 2. If sex is female and site is C60._,C61._,C62._, or C63._ 																																				
Sex/histology	<p>Some sex/histology combinations are unlikely:</p> <ol style="list-style-type: none"> 1. If sex is male and histological family is endometrial, placental, ovarian tumours 2. If sex is female and morphology is 9061-3; 9102 																																				
Behaviour/site	Behaviour code /2 is considered as unlikely by the program with the following sites: C40._; C41._;C42._;C47._; C49._; C70._; C71._; C72._																																				
Behaviour/histology	The combinations that are not listed in the morphology numeric list of ICD-O-3 are considered to be unlikely.																																				

Basis of diagnosis/histology	<p>The ICD-O-3 morphological code is NOT allocated for the purpose of specifying the basis of diagnosis. However, it would be unlikely for some specific morphological diagnoses to have been made without a histological examination. Certain combinations are exceptions to this general rule and are validated by the program.</p> <p>A non-microscopically confirmed diagnosis (basis of diagnosis code < 5) is accepted only with the following histological codes:</p> <ul style="list-style-type: none"> - Neoplasm, NOS (8000) - Islet cell tumours, gastrinomas (8150-8154) - Hepatocarcinoma (8170) - Pituitary tumours (8270-8281) - Melanoma of the eye (8720 and site is C69._) - Melanoma of skin (8720 and site is C44._) - Sarcoma, NOS (8800) - Neuroblastoma, NOS (8960) - Choriocarcinoma, NOS (9100) - Kaposi sarcoma (9140) - Craniopharyngioma (9350) - Glioma (9380) - Subependymal giant cell astrocytoma (9384/1) - Neuroblastoma, NOS (9500) - Retinoblastoma, NOS (9510) - Meningioma (9530-9539) - Lymphoma, NOS (9590) - Multiple myeloma (9732) - Waldenstrom macroglobulinemia (9761) - Leukaemia, NOS (9800) <p>Otherwise, the combination is considered as unlikely.</p>
------------------------------	---

If any of these checks on the data fail, CanReg gives warnings at the time of data entry. When the IARC-CHECK program is run in batch mode, it produces:

- A. An output data file, which has the same layout as the input file but with symbol(s) and the new codes written at the end or replacing the original codes.
- B. A warning file, created in the same directory as the input file, with the same name, but with the extension .CHK. It contains records which have been written to the output file, but which should be checked.
- C. An error file, created in the same directory as the input file, with the same name the extension .ERR. The file contains all invalid combinations of items; these records are NOT included in the output file.

9.2 MEASURING COMPLETENESS OF REGISTRY DATA

The population-based registry aims to record all cancer cases occurring within its defined geographical area. It is therefore essential that all the data sources for the registry be covered completely. That is, case-finding and abstracting should include all hospitals within the catchment area of the registry. All data sources within these hospitals should likewise be covered in order to avoid under-reporting.

There are a number of methods that provide some indication of the completeness of a registry, but which do not actually quantify the number of cases missing. They include the following, discussed in more detail below:

- 9.2.1. Historic data methods
 - 9.2.1.1 Stability of incidence rates over time
 - 9.2.1.2 Comparison of incidence rates in different populations
 - 9.2.1.3 Shape of age-specific curves
 - 9.2.1.4 Incidence rates of childhood cancers
- 9.2.2. Mortality:Incidence ratios
- 9.2.3. Number of sources/notifications per case

Three methods are available to obtain a quantitative evaluation of the degree of completeness of registration:

- 9.2.4 Independent case ascertainment
- 9.2.5 Capture-recapture methods
- 9.2.6 Death certificate methods

9.2.1. Historic data methods

9.2.1.1 Stability of incidence rates over time

If the registration area remains constant, then the number of cases registered per year might be expected to show only small and gradual changes from one year to the next. Quite often, the numbers of registrations will be increasing over time (as the population covered gets bigger, and older), so that looking at rates of incidence may be more useful.

However, it is very useful to use CanReg to prepare tables and graphs of time trends, by cancer type, and year, using the “Number of cases in major diagnosis groups in single calendar year of observation” option (Table 9.7).

Editorial table 1: Number of cases registered per year by site, and a bar chart of the total number of cases registered per year; see the chapter text for more details

EREWHON (2003–2007)

SITE	2006	2007	2008	2009	2010	Total
Lip, oral cavity and pharynx (C00–14)	8 (1.1)	11 (1.5)	11 (4.3)	9 (1.8)	7 (1.3)	71 (1.3)
Digestive organs (C15–26)	126 (17.7)	100 (13.9)	43 (18.8)	130 (21.5)	105 (20.2)	945 (17.1)
Respiratory organs (C30–39)	11 (0.1)	3 (0.4)	3 (1.2)	8 (1.5)	7 (1.3)	24 (0.9)
Bone, cartilage, melanoma (C40–43)	14 (2.0)	17 (2.4)	10 (3.9)	15 (2.9)	10 (1.9)	146 (2.6)
Male genital (C60–63)	335 (51.1)	238 (59.0)	79 (31.0)	177 (34.8)	276 (53.1)	2675 (48.3)
Urinary organs (C64–68)	24 (3.4)	29 (4.0)	23 (7.8)	45 (8.8)	18 (3.5)	301 (5.4)
Eye, brain, thyroid etc. (C69–75)	47 (6.6)	32 (4.4)	11 (4.3)	26 (5.1)	17 (3.3)	266 (4.8)
Haematopoietic (C81–96)	20 (2.8)	22 (3.3)	21 (8.2)	30 (3.9)	14 (2.7)	241 (4.4)
Other and unspecified	63 (8.8)	26 (7.8)	40 (17.7)	62 (12.2)	42 (8.7)	241 (9.8)
All sites but skin (C00–96bC44)	24 (3.4)	20 (2.8)	12 (4.7)	37 (7.3)	21 (4.0)	302 (5.5)
All sites but skin (C00–96bC44)	712 (106.0)	720 (100.0)	255 (100.0)	506 (100.0)	520 (100.0)	5540 (100.0)

SITE	2006	2007	2008	2009	2010	Total
Lip, oral cavity and pharynx (C00–14)	10 (1.5)	7 (1.0)	3 (0.8)	6 (0.9)	5 (0.8)	57 (1.0)
Digestive organs (C15–26)	70 (16.5)	84 (11.7)	53 (13.6)	82 (12.0)	67 (11.5)	656 (11.4)
Respiratory organs (C30–39)	2 (0.3)	2 (0.3)	1 (0.3)	6 (0.9)	1 (0.2)	16 (0.3)
Bone, cartilage, melanoma (C40–43)	13 (1.9)	25 (3.2)	10 (2.6)	22 (3.2)	8 (1.3)	140 (2.6)
Breast (C50)	136 (27.9)	217 (30.3)	35 (9.0)	67 (14.2)	126 (15.8)	1291 (23.4)
Female genital (C51–58)	23 (5.4)	4 (1.1)	3 (7.7)	36 (5.3)	60 (5.4)	120 (5.7)
Urinary organs (C64–68)	234 (32.1)	220 (30.7)	166 (42.6)	285 (41.7)	238 (40.5)	2063 (33.7)
Eye, brain, thyroid etc. (C69–75)	24 (3.6)	26 (3.6)	11 (2.8)	30 (4.4)	20 (3.1)	218 (3.8)
Haematopoietic (C81–96)	25 (3.7)	38 (5.3)	32 (8.2)	33 (4.8)	26 (4.1)	322 (5.6)
Other and unspecified	46 (6.9)	37 (5.7)	28 (7.7)	46 (6.7)	36 (5.7)	387 (6.6)
All sites but skin (C00–96bC44)	34 (5.1)	19 (2.6)	21 (5.4)	41 (6.0)	30 (4.7)	287 (5.0)
All sites but skin (C00–96bC44)	667 (100.0)	717 (100.0)	390 (100.0)	684 (100.0)	637 (100.0)	3774 (100.0)

SITE	2006	2007	2008	2009	2010	Total
Lip, oral cavity and pharynx (C00–14)	18 (1.3)	18 (1.3)	14 (7.7)	15 (1.3)	17 (1.0)	178 (1.1)
Digestive organs (C15–26)	196 (14.2)	184 (12.8)	101 (15.7)	152 (16.1)	172 (14.9)	1601 (14.2)
Respiratory organs (C30–39)	3 (0.2)	3 (0.3)	4 (0.6)	14 (1.2)	8 (1.7)	71 (0.6)
Bone, cartilage, melanoma (C40–43)	27 (2.0)	40 (2.8)	20 (3.1)	37 (3.1)	18 (1.6)	292 (2.6)
Breast (C50)	573 (41.4)	645 (44.9)	114 (17.7)	274 (22.0)	402 (34.7)	3966 (35.1)
Female genital (C51–58)	23 (1.7)	1 (0.1)	30 (4.7)	36 (3.0)	60 (5.2)	120 (1.9)
Male genital (C60–63)	234 (17.0)	290 (15.3)	166 (25.7)	285 (23.4)	238 (20.3)	2063 (18.7)
Urinary organs (C64–68)	24 (1.7)	29 (2.0)	20 (3.1)	45 (3.8)	18 (1.6)	301 (2.7)
Eye, brain, thyroid etc. (C69–75)	71 (5.1)	58 (4.0)	22 (3.4)	56 (4.7)	37 (3.2)	484 (4.3)
Haematopoietic (C81–96)	45 (3.3)	62 (4.3)	33 (8.2)	53 (4.4)	40 (3.5)	368 (3.0)
Other and unspecified	109 (7.9)	93 (6.5)	68 (10.5)	108 (9.1)	81 (7.0)	927 (8.2)
All sites but skin (C00–96bC44)	58 (4.2)	39 (2.7)	33 (5.1)	78 (6.5)	51 (4.4)	589 (5.2)
All sites but skin (C00–96bC44)	1379 (100.0)	1427 (100.0)	645 (100.0)	1193 (100.0)	1157 (100.0)	11344 (100.0)

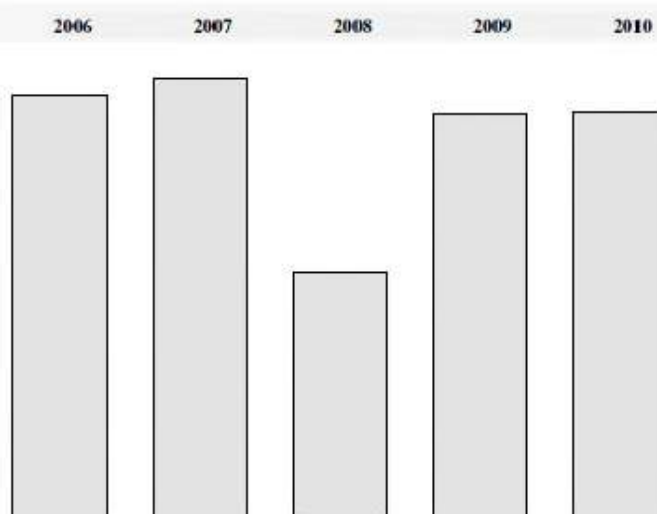


Table 9. 7 Number of cases registered per year by site, and a bar chart of the total number of cases registered per year (CanReg Table “Number of cases in major diagnosis groups in single calendar year of observation”)

Irregular numbers of cases suggest that case-finding was imperfect for the periods concerned. In the example shown, there is a falloff in the numbers of cases registered in 2008, and, although this affects many cancer types, the numbers of Kaposi sarcoma cases, in particular, shows a dramatic fall in that year. Under-reporting may be site specific e.g. researchers may have carried out some study on a particular cancer and medical records on patients involved may be taken out by the researcher and the registry staff may not locate them.

CanReg5 also allows time trends of the incidence rates (age standardised) of the major cancers to be plotted as a graph (Fig 9.1).

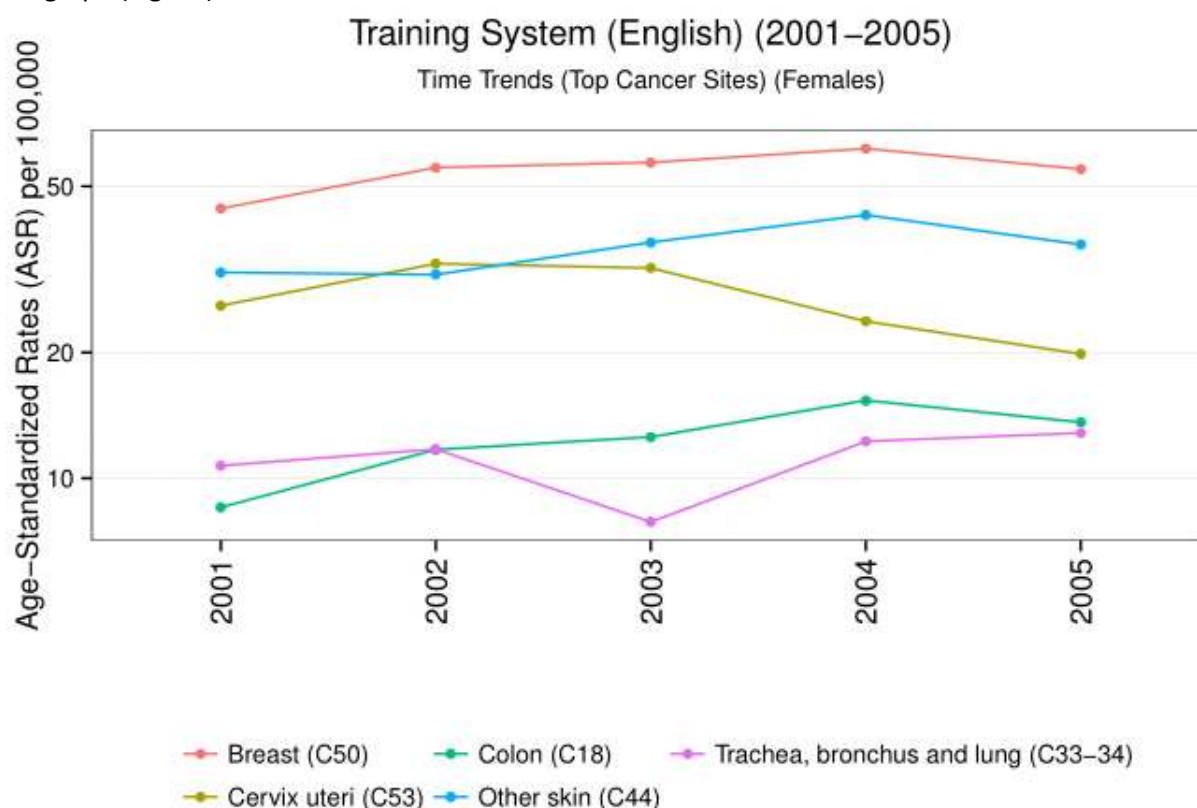


Fig 9. 1 Standard CanReg-5 output “Time trends (Top Cancer Sites)”

Another important check of registrations over time is to analyse the number of reports from different sources, by year of reporting.

In CanReg, a single tumour can be recorded as being found in several different sources. However, at the time of publication, there was no standard analysis within CanReg to calculate the number of notifications of a single case from different sources.

Procedure:

Make an export file of cases for the group of interest – defined by the years (for which the cases were registered) and the place of residence of the cases (geographic area).

For each case, make sure that the date of incidence and the codes for all of the different SOURCES of information are present.

Table 9.8 (left side) shows an example of an export file. For each source code (for example, hospital, or laboratory), see if it is present in Source 1, or Source 2, or Source 3, or Source 4 etc for all of the cases in the file.

Calculate the numbers of notifications from that source in one year. Table 9.8 (right side) shows an example of the calculation, using EXCEL.

Table 9. 8 Calculating no. of sources per registration: Example of an Excel Export File

Table 9. 9 shows an example of a table of results. This is a very useful check on reporting from the different sources, and suggests where case finding might have been incomplete at certain periods.

UASIN GISHU CASES: Number of sources															
Source	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	Total
00 DEATH CERTIFICATE	9	34	19	29	48	57	28	85	97	97	114	48	0	0	665
01 MR&TH	176	337	318	255	198	273	125	253	214	418	548	494	181	610	4400
02 ELDORET	11	14	17	16	22	12	12	10	2	4	7	3	2	1	133
03 PACIFICA	3	4	0	0	2	0	0	0	0	0	2	0	0	0	11
04 U.G MEMORIAL*	35	27	20	12	12	1	12	0	0	0	0	0	0	0	119
05 Kenyatta National	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
06 M.P. Shah	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
07 NAIROBI	1	2	0	0	0	0	1	1	1	0	1	0	1	0	8
10 Private Clinics	13	9	2	2	15	15	13	10	5	7	9	6	1	2	109
11 Private labs	0	0	0	2	0	0	5	21	1	2	3	3	0	1	38
12 HOSPICE	104	76	33	38	25	39	0	0	0	0	0	0	0	0	315
13 ELGON VIEW HOSPITAL	0	3	6	8	10	3	6	3	0	0	2	2	0	0	43
14 ITEN	0	0	2	0	1	0	0	0	0	0	0	0	0	1	4
15 KAPSABET	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
16 KITALE/MT.ELGON	0	0	3	0	0	3	15	0	0	0	0	0	0	2	23
17 KAPSOWAR	0	0	0	1	1	0	0	0	0	0	0	0	0	0	2
18 KAPEGURI	1	0	0	0	0	0	0	0	0	0	0	0	0	3	4
25 PLATEAU	1	0	0	0	3	0	0	0	0	0	1	2	0	0	7
50 HISTOLOGY LABS	58	217	123	159	159	213	126	210	59	266	328	377	424	751	3470
51 HAEMATOLOGY	5	14	33	32	67	78	46	26	50	11	12	40	170	113	697
60 RADIOLOGY	21	11	37	15	11	0	0	0	0	0	0	0	0	0	95
all sources	438	750	614	569	574	694	393	623	429	805	1030	978	779	1488	10164
cases	353	487	409	343	331	366	233	350	241	407	500	518	589	845	5972
average no. of sources per case	1.24	1.54	1.50	1.66	1.73	1.90	1.69	1.78	1.78	1.98	2.06	1.89	1.32	1.76	1.70

Table 9. 9 Analysis of the number of notifications from each source, by year (Eldoret Cancer Registry)

9.2.1.2 Comparison of incidence rates in different populations

The possibility of incomplete registration can be investigated by comparing observed incidence rates with expected values, based on those observed in registries in the same region, or estimated for the country in the latest edition of GLOBOCAN. The assumption is that the incidence rates for specific cancers should be rather similar to those from elsewhere in the same region (or country).

A standard Table (Data Quality Indicators) similar to that used by the editors of CI5 (Table 9.10) is planned for the "Table builder" option of CanReg 5 for this purpose.

This table presents the age-standardized incidence rates (and their standard errors) for 21 sites (and the total for all sites) in males and females, along with the ratio of the observed value to the expected value (O/E).

If the observed age standardized rate is significantly different from the expected value for the corresponding country or region, the O/E is shown in bold and flagged with a greater-than symbol (>) if the value is higher than expected or a less-than symbol (<) if the value is lower than expected.

EREWYHON (2003–2007)							
MALE							
SITE	Cases	ASR (se)	O/E	MV(%)	DCO(%)	M/I(%)	ICD-10
Lip, oral cavity and pharynx	6968	14.5 (0.18)	0.98	91.2	7.2	38.9	C00–14
Oesophagus	3293	6.3 (0.11)	0.95	83.4	14.7	75.4	C15
Stomach	7481	12.5 (0.15) <	0.83	76.9	19.5	61.2	C16
Colon, rectum and anus	27365	47.3 (0.30)	1.09	85.6	13.1	38.4	C18–21
Liver	4185	7.4 (0.12) >	1.26	58.0	37.3	82.9	C22
Pancreas	5339	9.3 (0.13)	1.03	56.0	35.7	91.4	C25
Larynx	2280	4.4 (0.10) <	0.88	87.0	12.0	39.3	C32
Lung (incl. trachea)	20052	35.2 (0.26) <	0.67	71.7 <	25.1	83.0	C33–34
Melanoma of skin	5703	11.5 (0.16) >	1.32	94.7	3.7	20.8	C43
Prostate	46799	79.2 (0.38) >	1.33	86.3	11.1	17.8 <	C61
Testis	2922	8.5 (0.17)	1.07	92.9 <	2.1	3.6	C62
Kidney etc.	7345	13.4 (0.17)	0.92	84.8	13.1	41.4	C64–66
Bladder	12303	20.5 (0.19)	0.98	91.2	7.6	16.5 <	C67
Brain, central nervous system	2777	6.3 (0.13)	1.04	73.7	21.6	70.5	C70–72
Thyroid	1493	3.3 (0.09) >	1.70	93.5	5.0	14.8 <	C73
Lymphoma	7747	15.4 (0.19)	1.02	81.7 <	16.3	43.5	C81–88,C90
Leukaemia	4619	9.5 (0.17) <	0.88	70.3	29.0	62.0	C91–95
Ill-defined (2.2% of total)	3941	6.8 (0.11)	0.87	50.7	43.4	132.9	C76–80
All sites but non-melanoma skin	179172	324.2 (0.81)	1.02	81.6	15.8	43.7	C00–96bC44
FEMALE							
SITE	Cases	ASR (se)	O/E	MV(%)	DCO(%)	M/I(%)	ICD-10
Lip, oral cavity and pharynx	2076	3.7 (0.09)	1.04	89.5	8.8	33.4	C00–14
Oesophagus	741	1.1 (0.04)	0.84	76.2	22.0	75.6	C15
Stomach	6017	6.9 (0.11)	0.90	70.7	25.2	67.0	C16
Colon, rectum and anus	22635	28.1 (0.22)	1.02	80.4	18.1	41.7	C18–21
Liver	1621	2.1 (0.06)	1.02	48.5	46.7	93.8	C22
Pancreas	5433	6.2 (0.10)	1.01	44.8	47.5	93.6	C25
Larynx	274	0.5 (0.03)	0.85	84.7	14.2	40.5	C32
Lung (incl. trachea)	8461	13.1 (0.16)	0.95	72.2	24.3	78.4	C33–34
Melanoma of skin	5795	11.3 (0.16) >	1.27	94.8	3.8	15.3	C43
Breast	48551	82.6 (0.41) >	1.15	86.9 <	11.2	26.9	C50
Cervix uteri	3523	7.1 (0.13) <	0.76	89.5	8.4	31.8	C53
O&U part of uterus	8689	13.0 (0.15)	1.03	88.8 <	10.0	23.3	C54–55
Ovary	6672	10.4 (0.14)	1.01	75.2 <	21.8	64.9	C56
Kidney etc.	4702	6.6 (0.11)	0.89	79.9	17.6	40.9	C64–66
Bladder	4315	5.2 (0.09)	1.04	86.3	12.3	25.0 <	C67
Brain, central nervous system	2405	4.6 (0.11)	1.02	66.9	28.2	67.2	C70–72
Thyroid	3778	8.4 (0.15) >	1.68	94.3	4.5	10.7 <	C73
Lymphoma	7156	11.3 (0.16)	1.04	77.8 <	20.2	46.7	C81–88,C90
Leukaemia	3743	6.1 (0.13)	0.90	62.7	36.8	68.6	C91–95
Ill-defined (2.7% of total)	4228	4.5 (0.08)	0.89	38.9	55.1	114.0	C76–80
All sites but non-melanoma skin	158408	243.8 (0.71) >	1.05	79.4	18.1	44.0	C00–96bC44

Data compared with those from seven registries in the same region/country.
Significantly lower (<) or higher (>) values are shown in bold.

Table 9. 10 Data Quality Indicators: Standard Table

9.2.1.3 Shape of age-specific curves

CanReg 5 produces a set of graphs showing age-specific incidence for 12 cancer sites (one curve for each sex) – option “Age-specific rates for major/most diagnosis groups (semi-logarithmic)” in Table builder (Figure 9.2).

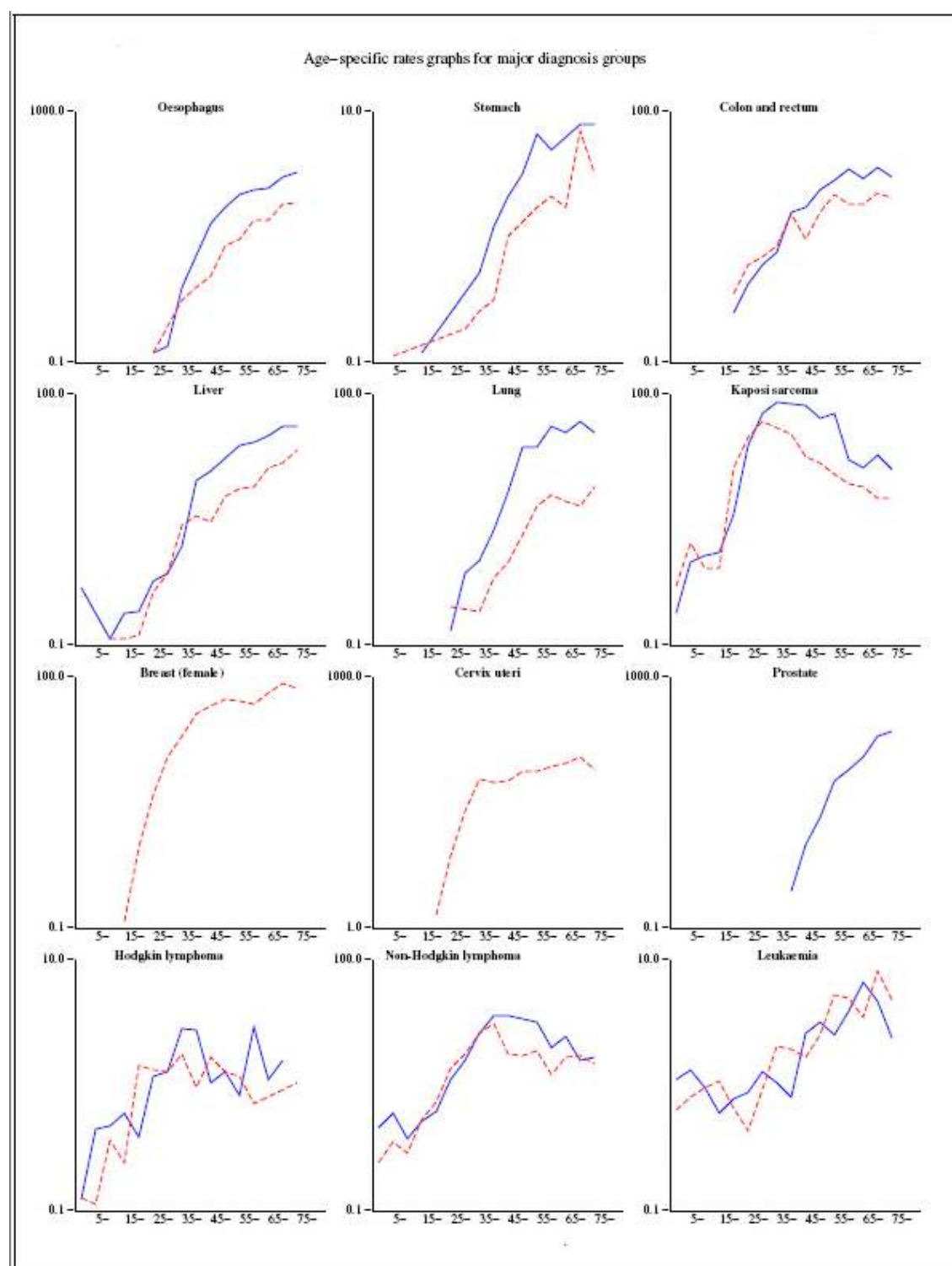


Fig 9. 2 Age-specific rates graphs for major diagnosis groups

The curves can be examined in order to detect abnormal fluctuations in the anticipated patterns, including any fall-off in the rate of increase in incidence in older subjects (which may be indicative of under ascertainment within these groups (although there can also be other explanations).

9.2.1.3 *Incidence rates of childhood cancers*

With respect to childhood cancer, the incidence rates (for all types combined) in the childhood age groups (0-4, 5-9, and 10-14) show much less variability than in adults. The possibility of under-enumeration (or duplicate registrations) in this age range can be investigated by comparing the observed age-specific rates in the childhood age range with an “expected” range of values.

The lowest and highest deciles of incidence rates of childhood cancer in the CI5 Volume IX data are shown in Table 9. 11.

Age group (years)	Boys		Girls	
	Lowest	Highest	Lowest	Highest
0-4	<13.7	>25.6	<11.3	>23.3
5-9	<8.9	>16.5	<7.0	>23.2
10-14	<9.2	>16.3	<7.9	>14.9

Table 9. 11 The lowest and highest deciles of incidence rates (per million) of childhood cancer in Volume IX

9.2.2. Mortality:Incidence Ratios

The Mortality:Incidence (M:I) ratio is an important indicator of completeness for cancer registries. It compares the number of deaths, obtained from a source independent of the registry (usually, the vital statistics system), and the number of new cases of a specific cancer registered, in the same period of time.

This method cannot be used where there is no comprehensive death registration, or when cause of death is missing or inaccurate on death certificates, which is the situation in almost all countries in Africa.

For the very few countries with reasonable death registration, the M:I ratios can be compared to standard values from the same region, testing for significant differences (see Table 9.10). When the quality of the mortality data is good and incidence and survival are in steady state, the M:I ratio is approximately 1 minus the 5-year survival probability (Asadzadeh Vostakolaei et al, 2011).

M:I ratios that are higher than expected raise suspicion of incompleteness (i.e. incident cancers missed by the registry), especially if the values are high for several different sites. However, under- or over reporting of tumours on the death certificates distorts this relationship.

9.2.3. Number of sources/notifications per case

The reason for using as many sources of information as possible on cancer cases, is that it reduces the likelihood of missing cases. Multiple sources reporting thus increase the completeness of the registry data.

When examining the number of reports from different sources, by year of reporting (see section 9.2.1.1) it is simple to compare the total number of sources reporting cases in one year, with the number of cases registered (See Table 9. 9). The ratio is the average number of sources per case.

In the example shown (Table 9.9) it is clear that the average is low (1.3) in one year (2010), compared to the average for the whole 14 year period (1.7 sources per case) due to poor reporting from two sources (MR& TH and Death Certificates) in that year.

9.2.4 Independent case ascertainment

Two methods can be used:

- Re-screening the sources that had been used by the registry, to detect any case missed during the registration process (Case finding audits);
- Using independent sources of cancer cases (which have not been used by the registry), and comparison of the registry database with them.

9.2.4.1 Case finding audit

Here, the idea is to go back to one or more of the registries data sources, and do an independent case finding (and abstracting) exercise. NAACCR suggests that each source (“reporting facility”) should be routinely audited at least once every 3 years (Hofferkamp, 2008). Audits also should be conducted when there is a documented decline in case reports from a facility (compared with the numbers of reports in the previous year’s – see Section 9.2.1.1).

Records of cancer cases identified during the audit are enumerated and matched against the registry’s files. Unmatched cases are followed back to verify whether they meet the reportability criteria (Section 3.2). The percentage of cases actually missed that *should* have been reported is calculated.

Most such studies focus on hospital sources. They thus provide an estimate of the completeness of reporting for those sources, not a true estimate of completeness for the whole registry (which is using multi-source reporting).

9.2.4.2 Using an independent source

Here, we need to find a list of cancer cases that have been compiled independently of the cancer registry’s case-finding procedures.

Comparing this list of cases with the registry database is a particularly useful method of evaluating completeness.

It requires record linkage between the cancer registry database and the independent case series, to estimate the numbers of cases in the latter “missed” by the registry.

Record linkage can be done using the “REC-LINK” software, originally developed by IARC⁶ or other more recently developed free record linkage software such as LinkPlus:

<http://www.cdc.gov/cancer/npcr/tools/registryplus/lp.htm>

or FRIL :

<http://fril.sourceforge.net/>

The proportion of eligible patients who are already registered is a direct and quantitative estimate of completeness. What independent sources of cancer cases might exist?

The most usual sources in Africa are:

- Cases recruited into local clinical trials
- Cases recruited into special studies (e.g. hospital case control studies)
- Cases recorded in databases by individual clinicians
- Cancer deaths from special community studies (e.g. verbal autopsy studies)

⁶ REC-LINK is available from the AFCRN secretariat

The text box below describes a study of completeness of registration in the Kampala Cancer Registry, using an independent data source (a case-control study of HIV and cancer), which recruited cancer cases from the main hospital, using project staff who worked independently of the cancer registry (Parkin et al, 2001).

THE KAMPALA COMPLETENESS STUDY

Record linkage and verification

A file was prepared of cases of cancer enrolled into the HIV Cancer study, with a date of diagnosis during 1994-1996 and resident in Kyadondo County. This file was linked to the master file of the cancer registry in June 1998 (approximately 18 months after the recorded diagnosis of the last case).

Record linkage was performed using the program REC-LINK. This provides a probabilistic matching of files based upon selected variables; in the current study, we used family name, first name, age, sex, and tribe. Each variable is assigned a score from 1 to 5 for reliability (reflecting the probability that individuals who are truly identical will have identical records with respect to the variable) and discriminating power (reflecting the probability that identical records with respect to the variable truly belong to the same individual). The product of the reliability score and the discriminating powers score gives a weighting factor for that variable.

When two records are compared, the two values for each variable are compared to arrive at a Similarity score. For each variable, its weighting is multiplied by this similarity score, and the sum (expressed as a percentage) gives the final probability that these two records belong to the same person.

Possible linkages (with percentage scores between 70 and 90) were checked "manually", using additional recorded variables (including maiden name, address, diagnosis, hospital, laboratory number) to make the final decision.

Cases that could not be linked to the cancer registry were examined in more detail. The study questionnaires were reviewed and, for many of the subjects, so too were many cases, their inclusion in the data set for record linkage was an error. This was the result of mistakes in recording of information, or in data coding and data entry, especially with respect to date of diagnosis (several prevalent cancer cases had been enrolled into the study), diagnosis (some enrolled cases had proved not to have cancer when the full range of diagnostic tests was completed), and address. These cases were deleted from the file.

Estimation of completeness

Case-finding for the cancer registry was performed independently of recruitment into the HIV Cancer study and completeness of registration was therefore estimated from the percentage of eligible cases in the HIV Cancer study data set which had been included in the registry database. Ninety-five percent confidence limits were calculated by the "exact" method.

To investigate the independent contribution of different patient variables to completeness of ascertainment, we fitted a logistic regression model to the data using STATA. The outcome variable was detection (or not) of cases by the registry, and the explanatory variables were sex, age (continuous and grouped), year, basis of diagnosis (microscopic or not), and diagnosis. The initial age groups used were 15-29, 30-39, 40-49, 50-59, and 60-69. Diagnoses were initially considered as 23 cancers, and then regrouped into seven categories (Kaposi's sarcoma, cervix cancer, oesophagus, liver, breast, eye, and all other).

9.2.5 Capture-Recapture Methods

Capture-recapture methods are possible when cancer registries use two or more different types of source for case finding. For registries using CanReg (version 4 or 5), each cancer case registered in the database may have several sources (where the information on the case was found). For a capture-recapture analysis, the different sources need to be regrouped into two or three reasonably independent categories:

Source 1: Hospital Source (hospital records departments/ radiotherapy/ oncology etc)

Source 2: Pathology Labs (include cytology, haematology)

Source 3: Death Certificates

For capture-recapture estimates of completeness to be accurate, identification (capture) of a case by one type of source should be independent of the other(s). In practice, this is a bit unlikely. For example, cases of cancer admitted to hospital might be more likely to be found in the pathology laboratory records labs than cases that were not. Or patients who die (and get a Death Certificate) may be less likely to be found from pathology (if they died soon after reaching hospital), than patients who do not die.

Although these limitations may cause estimates of completeness to be a bit inaccurate, in practice they are not too bad (Parkin et al, 1994).

Estimate with TWO types of source

When only two types of source are available (usually this will be hospital and pathology sources), cases registered in a given period (usually one year) can be classified as being registered from information obtained from one, or the other, or both, as in Table 9.12.

		SOURCE A	
		PRESENT	ABSENT
SOURCE B	PRESENT	a	b
	ABSENT	c	d

Table 9. 12 Estimate of Completeness by Capture-Recapture using TWO types of source

The number of cases from neither source (*d*) represent cases “missed” by case finding.

d is estimated as **bc/a**

and completeness will be $\frac{(a+b+c)}{(a+b+c+d)}$

Estimate with THREE types of source

If registrations can be classified by all three types of source, then 3 estimates of completeness can be made, and the true values should be somewhere between the highest and lowest values. The method is to use the above “paired” estimate, using, for each source, the combination of the other two as the other half of the pair. Using the example of Hospital (HOSP), Pathology (PATH) and Death Certificate (D.C.) as three sources, there are 7 possible combinations of sources (Fig 9. 3), and three sets of paired tables can be constructed (Table 9. 13).

For each table, we estimate the number of cases d that appear in none of the three sources.

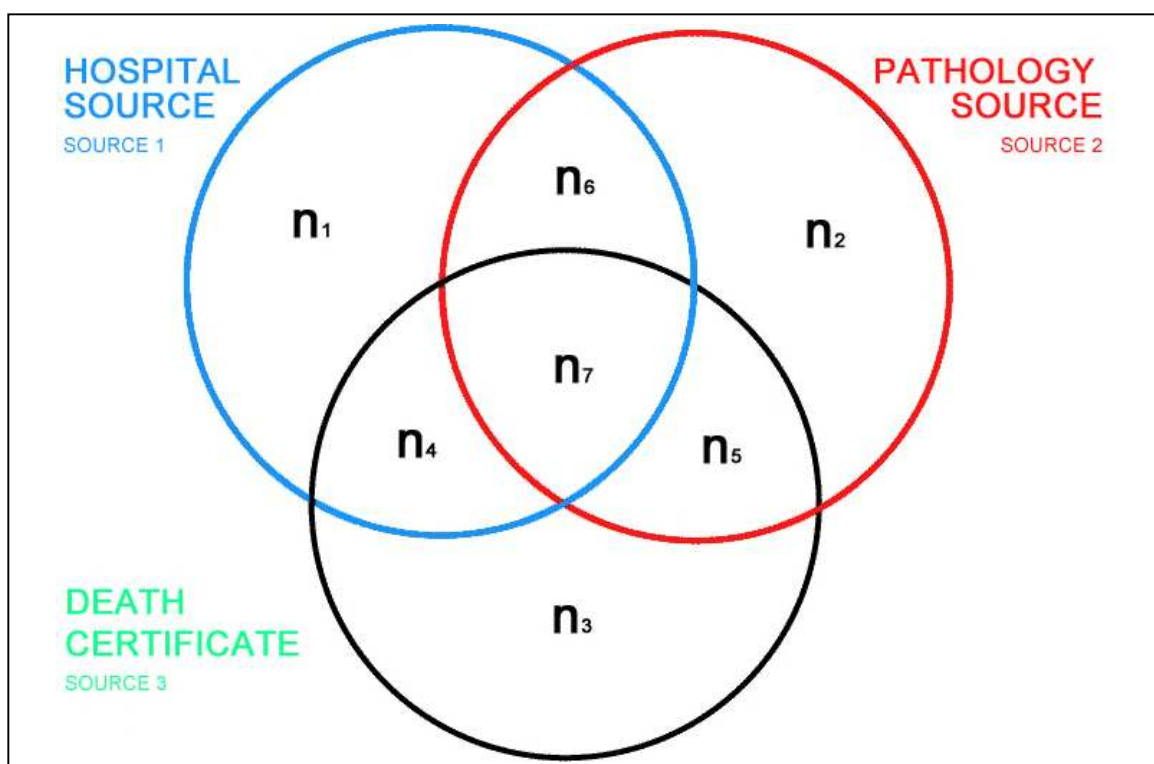


Fig 9. 3 Possible Combinations of THREE Sources

		HOSPITAL	
		PRESENT	ABSENT
PATH. and/or D.C.	PRESENT	$n_6 + n_7 + n_4$	$n_2 + n_5 + n_3$
	ABSENT	n_1	d

		PATHOLOGY	
		PRESENT	ABSENT
HOSP and/or D.C.	PRESENT	$n_6 + n_7 + n_5$	$n_1 + n_4 + n_3$
	ABSENT	n_2	d

		Death Certificate	
		PRESENT	ABSENT
PATH and/or HOSP.	PRESENT	$n_4 + n_7 + n_5$	$n_1 + n_6 + n_2$
	ABSENT	n_3	d

Table 9. 13 Estimate of Completeness by Capture-Recapture using THREE types of sources

Differences between the estimates are due to the interdependence of pairs of sources (clinical records, pathology, death certificates).

Various more sophisticated methods are available for dealing with the problem of dependency between sources. Larsen et al (2004) identified the degree of dependence between pairs of sources, then grouped those sources with the most dependence, before estimating the missing cases in a two way method with a third source. When the sources are all dependent, this approach cannot work, and log-linear modelling is needed (Parkin & Bray, 2009).

9.2.6 Death Certificate Methods

Using death certificates as a source of information can provide a very useful method of evaluating completeness. If the registry is finding many new cases via death certificates, it is certain that registration is incomplete (since some patients who do NOT die will have been missed by case finding).

To estimate completeness, we need to know the number of cases which come to the attention of the registry for the first time through a death certificate (sometimes called “Death Certificate Notifications” (DCN)). These are not the same as Death Certificate Only (DCO) cases – the latter are only a fraction of the cases which are first identified from a death certificate (those that cannot be traced by follow-back to the source of the death certificate (see Section 6. DEATH CERTIFICATE NOTIFICATIONS, page 33)). An estimate of the degree of completeness also needs the ratio of cases to deaths (M:I ratio – see section 9.2.2, page 58) (Parkin & Bray, 2009).

African registries can rarely calculate the number of DCN cases, and the M:I ratio can only be calculated when good quality vital registration of deaths, by cause, is present – the case for very few cancer registries. Death Certificate Methods can very rarely be used, therefore.

The competitions on limited resources in Africa due to other health problems that include HIV/AIDS make cancer registration sometimes a dream hard to realise. For those registries managed in whatever form; do not give up, keep going and remember that with imagination, patience, perseverance, dedication and willingness to solve local problems; establish population-based cancer registration is possible.