Catalogue no. 82-225-XIE — No. 006
ISSN: 1715-2100
ISBN: 0-662-41813-1

Canadian Cancer Registry Manuals

User Guide to Data Quality Reports for Provincial/Territorial Cancer Registries

by Michel Cormier

Health Statistics Division
Client Custom Services
Room 2200, Main Building, Ottawa, K1A 0T6

Telephone: 1 613 951-1746

Statistics Canada Statistique Canada

Canada

Canadian Cancer Registry Manuals

User Guide to Data Quality Reports for Provincial/Territorial Cancer Registries

by

Michel Cormier

82-225-XIE No. 006
ISSN: 1715-2100
ISBN: 0-662-41813-1

Health Statistics Division
Room 2200, Main Building, Ottawa, K1A 0T6

How to obtain more information :
Client Custom Services Unit: 1 613 951-1746
E-Mail: HD-DS@statcan.ca

October 2005

Published by authority of the Minister responsible for Statistics Canada

La version française de cette publication est disponible sur demande (n° 82-225-XIF au catalogue).

Table of contents

Page

1.0 Introduction

The purpose of the Canadian Cancer Registry (CCR) Data Quality Reports is to provide detailed feedback on the quality of data submitted each year through the CCR Core Edit system. The system is maintained by the Operations and Integration Division (OID) of Statistics Canada. There are two main parts to the reports. Part one presents data quality indicator percentages on all invasive tumour records included in a given data year. It also includes the number of breast tumours and the number of deaths reported in each data year. The second part provides information on the percentage of missing values for all invasive tumours and for all patients, as well as the number of deaths that are reported for each year.

These reports are produced at least twice per year, prior to the Canadian Council of Cancer Registries (CCCR) Data Quality Committee meetings. The reports are run from the CCR Tabulation Master File (TMF) created for the Health Statistics Division of Statistics Canada.

These reports are intended for use by the Provincial/Territorial Cancer Registries (PTCRs) to monitor the quality of their information on the CCR database. Potential errors may be identified in situations where the information provided meets all edit requirements, but lacks in completeness or accuracy. The reports contain information on the individual PTCRs, as well as for all of Canada.

2.0 Background

These tables were developed at Statistics Canada, in consultation with the Data Quality Committee (DQC) of the CCCR. The tables contains a set of guidelines developed and approved by the CCCR, to assist the PTCRs in monitoring data quality (see section 4.0 of this report). Review of data in these reports, along with the guidelines, is intended to aid registries in identifying priorities for improvement. This review should, thereby, contribute to improvements in the Canadian system of cancer registration as a whole. Better knowledge of data quality will also aid in documenting data limitations in the provincial/territorial and national cancer registries.

To the extent that each PTCR's data meet or exceed the guidelines, better results can be expected for the ICD conversions in the Core Edit cycle, for identifying duplicates in the Internal Record Linkage cycle and for finding the correct death record for each patient in the Death Clearance cycle.

Adherence to these guidelines will help all PTCRs to ensure their data will meet standards for acceptance in international publications such as *Cancer Incidence in Five Continents*, *and Cancer Incidence in North America*.

3.0 Reports format

3.1 Columns

The first column indicates the name of the variable being reviewed on the table. The next columns indicate the PTCR name for one year of data at a time or the data year for one particular PTCR. The last two columns indicate, for the given variable, the CCR and SEER recommended guidelines. Some guidelines only list a range of values, which is explained below.

Examples of how ranges may be coded:

<5%: values should be below 5%
2-5%: values should lie within the range of 2 to 5%
[0(1-3)5]: values should, ideally, lie within the range of 1 to 3%, however, a value between 0 to 5% is acceptable

3.2 Footer

The name of the TMF that was used to create the reports is given at the bottom of each page.

3.3 Part 1: Data quality indicator percentages

The purpose of this part is to provide an in-depth look at selected indicators of the quality of diagnostic information for each tumour. This table is produced for all invasive tumours for one year of diagnosis. The basis for tabulation of each row of the report is described below. Field Numbers are described in the CCR Input Data Dictionary.

Both the **numerator and denominator of the percentage calculation include only the cases that are invasive** (ICD-O-2 - Behaviour = 3). The first row of this part of the report lists the total number of tumours for one particular year or registry.

Non-Specific ICD-O-2 - T Codes: This percentage is calculated as the number of records having the following codes for ICD-O-2 - Topography (Field T15) divided by the number of tumours on the file.

C76._ Other and Ill-defined Sites
C80.9 Unknown Primary Site

Recommended range: <5%

Unknown primary site (C80.9 only): This percentage is calculated as the number of records having only the ICD-O-2 - Topography (Field T15) code of C80.9 Unknown Primary site divided by the number of tumours on the file.

Non-specific ICD-O-2 - M codes: This percentage is calculated as the number of records with codes in the range of 8000-8004, Neoplasms, NOS, for ICD-O-2 - Morphology (Field T16), divided by the number of tumours in the file.

Recommended range: Ideal: 2-5%
Acceptable: 0-8%

Improbable combinations of site/morphology codes: This percentage is calculated as the number of records having an improbable code for ICD-O-2 - Topography (Field T15) for a given **specific** morphology for ICD-O-2 - Morphology (Field T17) divided by the number of tumours on the file. Appendix A (Part 1) contains the list indicating the site-specific morphologies.[1]

Recommended range: <1.0%

Registries should have a procedure in place to verify the accuracy of these records.

Improbable combinations of site/morphology/age codes: This percentage is calculated as the number of records having improbable combinations of ICD-O-2 - T (Field 15), of ICD-O-2 - M (Field 17) and of Age divided by the number of tumours on the file. Appendix A (Part 2) contains the list of these improbable combinations.

Recommended range: <0.02%

Method of diagnosis (Field T11): This set of indicators shows the percentage of records with each of the following values for this field.

- DCO (Death Certificate Only) (Field T11 = 6)

  Recommended range: Ideal: 1-3%
  Acceptable: 0-5%

- Microscopically confirmed, which includes:
  - Histology (Field T11 = 1)
  - Autopsy (Field T11 = 2)
  - Cytology (Field T11 = 3)
- Unknown (Field T11 = 9)

1. As developed by the CCCR Data Quality Committee, and accepted at the May 1994 meeting of the Canadian Couuncil of Cancer Registries in Halifax.

Recommended range: <2%

- Other diagnoses (Field T11 = 4, Radiology/laboratory or 5, Surgery or Clinical)

Histology/autopsy and ICD-O-2 - morphology = 8000: This combination shows the percentage of records where the Method of Diagnosis (Field T11) = 1 (Histology) or 2 (Autopsy) and ICD-O-2 - M (Field T16) = 8000. The combination of a histologically confirmed method of diagnosis with an unspecified morphology is unusual and all such cases should be reviewed by the PTCR.

Recommended range: <0.1%

Age at diagnosis: Age is calculated based on the Date of Birth (Field P11) and Date of Diagnosis (Field T12). Cases with age less than one year, or greater than or equal to 100 years of age should be reviewed, particularly if the percentage of such cases is outside the guidelines.

- <1 year: records where age is less than one year.
- >= 100 years: records where age is greater than or equal to 100.

Recommended range: <0.1%

Note: The formula used to calculate Age includes provisions for missing values, and is the same as the one used to derive Age at Diagnosis for the Tabulation Master File (TMF).

Male breast cancer: This indicator shows the percentage of male breast cancer cases where ICD-O-2 - T (Field T15) = C50._ and Sex (Field P10) = 1 (Male). The percentage is calculated as the number of male breast cancers divided by the number of all breast cancers occurring in both males and females. Cases coded to this combination should be reviewed, particularly if the percentage falls outside the guidelines. In ICD-O-2, the code for breast cancer does not distinguish between male or female. It is possible for records to be miscoded to male breast cancer as a result.

Recommended range: <0.7%

ICD-9 = 799.9: This indicator shows the percentage of cases where the ICD-9 code is 799.9 (Field P17 : Underlying Cause of Death is coded as “unknown” on the death certificate). This percentage is calculated as the number of **patients** with an ICD-9 code of 799.9 divided by the total number of deaths reported for that year.

3.4 Part 2: Percentage of missing values

The first section of Part 2 provides information on the percentage of **tumour** records on the Canadian Cancer Registry database that have missing or unknown information, for each variable, as described below.

Field #	Field Name	Unknown Value
T6	Place Name of Residence at Time of Diagnosis	All blank
T7	Postal Code	999999
T8	Coded Place of Residence at Time of Diagnosis: Census Division (CD) and Census Subdivision	CD unknown: PR00 CSD unknown: PRCD999
T10	Health Insurance Number	at least six "9s" or all "9s"
T12	Date of Diagnosis -month -day	Month unknown: YYYY9999 Day unknown: YYYYMM99
T19	Laterality - tabulated at % of applicable tumour records (laterality code of 1, 2, 4 or 9)	Unknown: 9

The **second** section of Part 2 provides information on all **patient** records on the Canadian Cancer Registry database that have missing or unknown information, for each variable, as described below.

Field #	Field Name	Unknown Value
P6	Current Surname	All blank (based on P5, Type of Current Surname=0, Surname Unknown)
P7	First Given Name: -% all blank -% just one letter	All blank (unknown or not applicable)
P10	Sex	Unknown: 9
P11	Date of Birth	Year unknown: 99999999 Month unknown: YYYY9999 Day unknown: YYYYMM99
P12	Province/Country of Birth	999 = Unknown
P13	Birth Surname (for female patients only)	All blank (Note: cannot be balnk if current surname is blank)

Items (P14 to P17) are calculated only for deceased patients, by excluding records for which fields P14 to P17 are zero-filled (i.e., N/A) as follows:

P14	Date of Death	Year unknown: 99999999 Month unknown: YYYY9999 Day unknown: YYYYMM99
P15	Province/Country of Death	Dead, unknown: 999 N/A (not known dead): 000
P16	Death Registration Number	Dead, # unknown: 999999 N/A (not known dead): 000000
P17	Underlying Cause of Death -COD (vital statistics) -COD (PTCR)	Unknown COD (VS): 7999 Unknown COD (PTCR): 0009 N/A: 0000

4.0 References

American Association of Central Cancer Registries. Standards for Cancer Registries Volume III. Standards for Completeness, Quality, Analysis and Management of Data. Seiffert JE (editor.) NAACCR Cancer Surveillance and Control Program. Sacramento, California, 1994.

Gao RN, Gaudette LA. Assessment of edits for improbable or impossible combinations of site, morphology and age for Canadian cancer registries. Health Statistics Division, Statistics Canada, Ottawa, 62 pp, August 1994.

Jensen OM, Parkin DM, MacLennan R, Muir CS, Skeet RG, editors. Cancer Registration Principles and Methods. Lyon: International Agency for Research on Cancer, IARC Scientific Publications No. 95, 1991:33,55.

Langlet ER and Gaudette LA. Data quality assessment for the National Cancer Incidence Reporting System, Canada and provinces 1961-1984. Health Statistics Division, Statistics Canada, Ottawa, 1989.

Parkin DM, Chen VW, Ferlay J, Galceran J, Storm HH, Whelan SL. Comparability and Quality Control in Cancer Registration. Lyon; International Agency for Research on Cancer. IARC Technical Report no. 19, 1994.

Parkin DM, Muir CS, Whelan SL, Gao YT, Ferlay J, Powell J, editors. Cancer Incidence in Five Continents, Volume VI. Lyon: International Agency for Research on Cancer, IARC Scientific Publications No. 120, 1992.

Appendix A: Improbable combinations requiring review

Part 1: Site-specific morphologies[1,2](Version: March 25, 1998)

Morphology Code	Site Code	Site Name
8147, 8941	C07._ C08._	Parotid Gland Other and unspecified major salivary gland
8142, 8144, 8145	C16._	Stomach
9764	C17._	Small intestine
8220	C18._	Colon
8240/1, 8243	18.1	Appendix
8123	C21.1	Anal canal
8124	C21.2	Cloacogenic zone
8170-8180, 8970, 9124	C22.0	Liver
8160, 8162	C22.1	Intrahepatic bile duct
8150-8152, 8154, **8452/1**, 8971	C25._	Pancreas
9521-9523	C30.0	Nasal cavity
8042-8045, **8140/1**, 8250-8251, 8972, **9134/1**	C34._	Bronchus and lung
8580	C37.9	Thymus
8812, 9180-9190, **9200/1, 9210/1** 9220-9230, 9250-9261, 9270-9330	C40._ C41._	Bone, joints and articular cartilage
9761	C42.0	Blood
9732, 9800-9910, 9931-9940, **9950/1, 9960/1, 9961/1, 9962/1, 9970, 9980/1, 9981/1, 9982/1, 9983/1, 9984/1, 9989/1**	C42.1	Bone marrow
8081/2, 8090-8095, 8110, 8247, 8390-8400, 8410-8420, 8721-8723, 8730-8761, 8780, 8832, 9700, 9709	C44._	Skin
9055/1	C48._	Retroperitoneum and peritoneum
8314-8315, **8500/3**, 8501-8503, 8512-8541, 8543, 9020	C50._	Breast
8076	C53._	Cervix uteri
8950	C54._	Corpus uteri
8930, **8931/1**	C54.1	Endometrium
8380-8381, 8441-8451, 8460-8473, 8600, 8620, **8621/1, 8622/1, 8623/1, 8632/1**, 9000, 9090, **9091/1**	C56.9	Ovary
9100/1, 9104/1	C58.9	Placenta
8080/2	C60._	Penis
8148/2	C61.9	Prostate gland
8640, 8650, 9061-9063, 9102	C62._	Testis
8312, 8361/1, 8960, 8964	C64.9	Kidney, NOS
8773-8774	C69._	Eye and adnexa
9510-9512	C69.2	Retina
9530, 9538/1, 9539	C70._	Meninges
9380-9382, 9383/1, 9384/1, 9391-9392, 9393/1, 9394/1, 9400-9460, 9473, 9481, 9594	C71._	Brain
9390	C71.5	Ventricle, NOS
9470-9472, 9480	C71.6	Cerebellum, NOS
8330-8350, 8511	C73.9	Thyroid gland

8370	C74.0	Cortex of adrenal gland

Appendix A: Improbable combinations requiring review

Part 1: Site-specific morphologies[1,2](Version: March 25, 1998)

Morphology Code	Site Code	Site Name
8700	C74.1	Medulla of adrenal gland
8322	C75.0	Parathyroid gland
8270-8281, 8300	C75.1	Pituitary gland
9350/1	C75.2	Craniopharyngeal duct
9360/1, 9361/1, 9362/1	C75.3	Pineal gland
8692/1	C75.4	Carotid body
8690/1, 8691/1	C75.5	Aortic body and other paraganglia

Notes:

1. For a given morphology, any record with a site **other** than those specified should be reviewed.

2. For a given morphology, the site applies to all behaviour codes (i.e., 0, /1, /2 or /3), except where the behaviour code (usually /1) is indicated. For example, records with morphology 8240 and a site other than C18.1 need only be reviewed where the behaviour code is /1.

Part 2: Age/site/morphology Combinations (For a given age and site, any record with a morphology listed below should be reviewed)

Age	Site	Morphology
< 010	Breast C50._	Any
	Vulva C51._	Any
	Vagina C52.9	Any
	Cervix uteri C53._	Any
	Corpus uteri C54._	Any
	Uterus, NOS C55.9	Any
	Placenta C58.9	Any
	Ovary C56.9	Any
	Other & unspecified female genital organs C57._	Any
	Other & unspecified male genital organs C63._	Any
	Bladder C67._	Any
	Other & unspecified urinary organs C68._	Any
	Thyroid gland C73.9	Any
< 020	Esophagus C15._	
	Small intestine C17._	Any
	Colon C18._	Any other than carcinoid 8240-8245
	Rectosigmoid C19.9	Any
	Rectum C20.9	Any
	Anus & anal canal C21._	Any
	Gallbladder C23.9	Any
	Other & unspecified parts of bilary tract C24._	Any
	Pancreas C25._	Any
	Trachea C33.9	Any other than carcinoid 8240-8245
	Lung and bronchus C34._	Any other than carcinoid 8240-8245
	Pleura C38.4	Any
< 030	Any	Multiple myeloma 9732
	Any	Chronic lymphocytic leukemia 9823
	Penis	Any
< 045	Prostate C61.9	Any
> 005	Eye & adnexa C69._	Retinoblastoma 9510-9512
> 014	Kidney C64.9	Wilms' tumor 8960
> 045	Placenta C58.9	Any