MINISTRY OF HEALTH
MANATŪ HAUORA

Statistics New Zealand
TATAURANGA AOTEAROA

IDI Data Dictionary:
Cancer registrations data

February 2017 edition

New Zealand Government

Liability
While all care and diligence has been used in processing, analysing, and extracting data and information in this publication, Statistics New Zealand gives no warranty it is error free and will not be liable for any loss or damage suffered by the use directly, or indirectly, of the information in this publication.

Citation
Statistics New Zealand (2017). *IDI Data Dictionary: Cancer registrations data (February 2017 edition)*. Available from www.stats.govt.nz.

ISSN 2537-7558 (online)

Published in February 2017 by
Statistics New Zealand
Tatauranga Aotearoa
Wellington, New Zealand

Contact
Statistics New Zealand Information Centre: info@stats.govt.nz
Phone toll-free 0508 525 525
Phone international +64 4 931 4600
www.stats.govt.nz

Contents

1 Purpose of this data dictionary 4
Background 4
2 About the Cancer registrations data 5
Overview 5
Coverage 5
Methodology 5
Quality information 5
Privacy, security, or confidentiality issues 6
3 Data dictionary for cancer registrations data 7
Dataset description 7
Detailed information 8

1 Purpose of this data dictionary

IDI Data Dictionary: Cancer registrations data (February 2017 edition) documents the content of this dataset that the Ministry of Health (MOH) provides to Statistics New Zealand to use in the Integrated Data Infrastructure (IDI).

This dictionary gives information on the variables contained in the dataset from 1995 – including technical information and descriptions.

Use this data dictionary if you are interested in understanding and accessing Cancer registrations dataset in the IDI for your research.

Background

The MOH seeks to improve, promote, and protect the health of New Zealanders. It does this through its sector leadership of New Zealand's health and disability system by:

- advising the Minister of Health, and government as a whole, on health issues
- directly purchasing a range of important national health and disability support services
- providing health sector information and payment services for the benefit of all New Zealanders.

The objectives of the MOH's data and metadata are to:

- measure and describe the information available within the National Collections
- promote uniformity, availability, and consistency across the National Collections
- support the use of nationally agreed protocols and standards wherever possible
- promote national standard definitions and make them available to users.

2 About the Cancer registrations data

Overview

The NZ Cancer Registry (NZCR) is a population-based tumour registry with a primary function to collect and store cancer incidence data that can then be used for cancer incidence and survival studies, public health research, monitoring screening programmes, and policy formulation.

See New Zealand Cancer Registry (NZCR) on the MOH website for more information, in particular, the technical details section for information on morphology coding changes.

See Cancer: New registrations and deaths – series on the MOH website for information about the annual series of' publications that report data from the NZCR.

Coverage

Reference period start: 1995

Reference period end: 2014

Geographic coverage: New Zealand

Target population: Healthcare users diagnosed with cancer in New Zealand

Observed population: People with a cancer registration

Methodology

Type of data: Administrative data capture

Data collector: NZ Cancer Registry (NZCR), National Collections and Reporting, MOH.

Mode of data collection: Some types/sites of cancer are of particular interest to researchers and the processing of these cancers is treated as a priority. These types/sites include: melanoma, lung, prostate, breast, cervix, colorectal childhood cancers, and malignant neoplasms of lymphoid haematopoietic tissue. The processing of data for these priority cancers is kept up to date to within three months of receipt of laboratory reports.

Frequency of data collection: Cancer registration data is loaded daily into the NZCR by a team of cancer coders. From here the NZCR datamart is refreshed daily from Monday to Friday.

Quality information

Other quality issues: Long-term trend data is unavailable for some fields and some cancer types. This may be due to increasing/decreasing field details and changes to the inclusion/exclusion criteria (eg blood and bladder cancers).

Multiple cancer events have been excluded. In other words, where a healthcare user has multiple cancer events at the same site, and with the same morphological type, only one has been deemed to be the primary cancer and included in this extract.

Privacy, security, or confidentiality issues

Privacy, security, and confidentiality issues are covered in chapters 2 and 5 of Microdata output guide.

The cancer registrations table accessible to researchers does not contain any name or address information to identify an individual. All researchers who have access to the cancer registrations data have had their research proposals assessed using Statistics NZ's microdata access protocols. Only approved researchers who have been granted access by Statistics NZ and the MOH may view the cancer registrations data.

Read Statistics NZ's microdata access protocols.

All outputs produced from cancer registrations data must be aggregated and counts suppressed if the underlying unrounded count is fewer than six.

3 Data dictionary for cancer registrations data

Dataset description

Contents of dataset: This dataset contains a subset of fields from the NZ Cancer Registry. Specifically it contains information on malignant cancer registrations for healthcare users in the population cohort.

IDI variable name	Primary key	Mand-atory	Format	Classification name	Variable name
snz_uid			Integer		
snz_moh_uid			Integer		
moh_can_event_id_nbr			Integer		Event ID
moh_can_birth_month_nbr			TinyInteger		
moh_can_birth_year_nbr			SmallInteger		
moh_can_domicile_code			Varchar, 4	domicile_code	Domicile code
moh_can_sex_code			Varchar, 1	gender_code	Sex
moh_can_ethnicity_1_code			char, 2	ethnic_code	Ethnicity 1
moh_can_ethnicity_2_code			char, 2	ethnic_code	Ethnicity 2
moh_can_ethnicity_3_code			char, 2	ethnic_code	Ethnicity 3
moh_can_eth_priority_grp_code			char, 2	ethnic_code	Prioritised ethnic group
moh_can_site_code			Varchar, 8		Site code
moh_can_morphology_code			Varchar, 8		Morphology code
moh_can_extent_of_disease_code			Char, 1	extent_of_disease_code	Extent of disease
moh_can_date_of_diagnosis_text			char, 10		Date of diagnosis
moh_can_basis_of_diagnsis_code			Integer	basis_of_diagnosis_code	Basis of diagnosis
moh_can_laboratory_code			Varchar, 4	laboratory_code	Laboratory code
moh_can_grade_tum_code			Varchar, 100		
moh_can_tnm_m			Varchar, 100		

IDI variable name	Primary key	Mand-atory	Format	Classification name	Variable name
moh_can_tnm_n			Varchar, 100		
moh_can_tnm_t			Varchar, 100		

Detailed information

IDI variable name: snz_uid

Definition: a global unique identifier created by Statistics NZ. There is a snz_uid for each distinct identity in the IDI. This identifier is changed and reassigned each refresh.

Format: Integer

Name of classification:

Notes:

IDI variable name: snz_moh_uid

Definition: a local unique identifier derived by Statistics NZ from the source agency's unique identifier(s). This identifier will remain the same for an identity across refreshes. Where we receive more information during a subsequent refresh that indicates that two or more identities represent the same identity, the identifier may change.

The snz_moh_uid represents a distinct identity in all of MoH tables in IDI.

Format: Integer

Name of classification:

Notes:

IDI variable name: moh_can_event_id_nbr

Definition: The unique identifier of the cancer event, assigned by the NZCR system.

Format: int

Name of classification:

Notes:

IDI variable name: moh_can_birth_month_nbr

Definition: The month on which the healthcare user was born.

Format: Tinyint

Name of classification:

Notes: Since September 2008, date of birth has been automatically populated from the National Health Indix (NHI) at the time a cancer event is created.

IDI variable name: moh_can_birth_year_nbr

Definition: The year on which the healthcare user was born.

Format: Smallint

Name of classification:

Notes: Since September 2008, date of birth has been automatically populated from the NHI at the time a cancer event is created.

IDI variable name: moh_can_domicile_code

Definition: A 4 digit code representing the healthcare user's usual residential address around the date of diagnosis.

Format: Varchar, 4

Name of classification: domicile_code

Notes: Since September 2008, domicile code has been automatically populated from the NHI at the time the cancer event is created. As such, the caveats around the domicile code in the NHI should also be considered here. For instance, it should be noted that before the NHI moved to its new platform in 2012, the address fields in the NHI were free-text with little validation. This meant there could be considerable variability in accuracy, which, in turn, meant addresses could not always be successfully geocoded to a domicile code, or could result in rural addresses being assigned to an urban domicile code where there was insufficient data to generate the correct code. This is because the automated geocoding software relies on generating a post code in order to determine where in a related table it should look to find the code. However, a number of validation checks were included when the NHI moved to its new platform and the quality of address information should improve markedly.

The domicile code is taken from either the 1991, 1996, 2001, 2006, or 2013 census domicile codes depending on the year of diagnosis. If the year of diagnosis was between:

- up to 31 December 1997, the 1991 code is used
- 1 January 1998 – 31 December 2002, the 1996 code was used
- 1 January 2003 – 31 December 2007, the 2001 code was used
- 1 January 2008 – 30 June 2015, the 2006 code was used
- 1 July 2015 onwards, the 2013 code was used.

For retrospective cancer registrations (.e, late notifications) the domicile code was assigned from the appropriate census code table relating to the year of diagnosis. In such cases, the domicile code may not truly represent the domicile at the time of diagnosis, as the address at the time of diagnosis may not be reported.

IDI variable name: moh_can_sex_code

Definition: The biological sex of the healthcare user.

Format: Varchar, 1

Name of classification: gender_code

Notes: The term ‘sex’ refers to the biological differences between males and females, while the term ‘gender’ refers to a person's social role (masculine or feminine).

Since September 2008, sex has been automatically populated at the time a cancer event is created. For cancer events created manually, a copy of the current sex from the NHI is used. For cancer events created from National Minimum Data Set (NMDS) or Mortality records, the sex from the source record is used.

IDI variable name: moh_can_ethnicity_1_code, moh_can_ethnicity_2_code, moh_can_ethnicity_3_code,

Definition: Ethnicity is the ethnic group or groups people identify with or feel they belong to. Thus, ethnicity is self-perceived and people can belong to more than one ethnic group.

Format: char, 2

Name of classification: ethnic_code

Notes: Ethnicity fields in the NZCR datamart are determined by running an algorithm across the ethnicity fields in the NHI, Mortality Collection and National Minimum Data Set (NMDS) to maximise the selection of meaningful ethnic codes and minimise the selection of default codes such as ‘Not Known’.

IDI variable name: moh_can_eth_priority_grp_code,

Definition: Ethnicity is the ethnic group or groups people identify with or feel they belong to. Thus, ethnicity is self-perceived and people can belong to more than one ethnic group. Where more than one ethnic group is reported, the Statistics NZ prioritisation algorithm is used to report only a single ethnicity.

Format: char, 2

Name of classification:

Notes: See National Minimum Data Set on the MOH website for information about the Statistics NZ prioritisation method and other aspects of the collection of ethnicity data at:

IDI variable name: moh_can_site_code

Definition: A code defining the site (or topography) of the tumour.

Format: Varchar, 8

Name of classification: site_list

Notes: The cancer site is coded using the International Statistical Classification of Diseases and Related Health Problems, Australian modification. The specific version used for each year of diagnosis is as follows:

- From Jan 1995 – Dec 1999 the version was ICD-9-CMA-II
- From Jan 2000 – Dec 2002 the version was ICD10-AM-II
- From Jan 2003 – Dec 2008 the version was ICD10-AM-III
- From Jan 2009- Dec 2013 the version was ICD10-AM-VI
- From Jan 2014 onwards the version is ICD10-AM-VIII

The malignant cancer site codes are in the range C00-C96. From 2003 onwards D45-D47 were considered malignant and were also recorded.

IDI variable name: moh_can_morphology_code

Definition: A code defining the morphology (histology, form, and structure) of the tumour.

Format: Varchar, 8

Name of classification: morph_list

Notes: Morphology is coded using the International Classification of Diseases for Oncology (ICD-O). Morphology data with a diagnosis year before 2003 has been coded in ICD-O 2nd edition, while data since that date has been coded in ICD-O 3rd edition.

IDI variable name: moh_can_extent_of_disease_code

Definition: A code describing the stage of development reached by the tumour at diagnosis

Format: Char, 1

Name of classification: extent_of_disease_code

Notes: Extent of disease is sourced from laboratory reports, hospital discharge events (NMDS) or mortality events (Mortality Register).

The extent of disease code is only provided from 1997 onwards.

IDI variable name: moh_can_date_of_diagnosis_text

Definition: The date the tumour was diagnosed.

Format: char, 10

Name of classification:

Notes: Sourced from laboratory reports, hospital discharge event start date (from NMDS), or mortality event date of death (from Mortality). This is the same as the earliest date of operation/biopsy or the date of admission (in a hospital event - NMDS) or the date of death if diagnosed on autopsy. If registering from a laboratory report, this should be the earliest laboratory report date. If the only notification of a cancer comes from a Medical Certificate of Causes of Death, the

diagnosis date is estimated from the 'Approximate time between onset and death' as reported by the certifying doctor alongside the cancer diagnosis on the certificate.

IDI variable name: moh_can_basis_of_diagnsis_code

Definition: A code used to describe the single, most valid basis of diagnosis for a primary malignant tumour, or a secondary tumour if the primary tumour site is unknown or cannot be determined.

Format: int

Name of classification: basis_of_diagnosis_code

Notes: Examples include: death certificate only, clinical only, histology of primary etc. The basis of diagnosis code is sourced from laboratory reports, or inferred from the source of the registration. In other words, the registration source code and basis of diagnosis code need to be consistent. For instance if the registration source code is ‘Laboratory source’, then the basis of diagnosis code might be ‘Histology of primary’ but is unlikely to be ‘Clinical only’.

IDI variable name: moh_can_laboratory_code

Definition: A code identifying the laboratory reporting the cancer event.

Format: Varchar, 4

Name of classification: laboratory_code (see Laboratory code table in the MOH website).

Notes:

IDI variable name: moh_can_grade_tum_code

Definition: A code that specifies the differentiation of the tumour.

Format:

Name of classification: Grade of tumour look-up table.

Notes: Introduced in 1998. Prior to this, general tumour differentiation information was held in the Morphology description.

IDI variable name: moh_can_tnm_m

Definition: An international classification code to describe the extent/size of the primary tumour, as classified by tumour size, nodes, and metastases (TNM), and is specific to the site. ITM is an international staging system comprising:

- pathological staging based on histology reports, and
- clinical staging based on hospital events.

Format:

Name of classification:

Notes: Refer to TNM Classification of Malignant Tumours for further details. Introduced in 2001.

IDI variable name: moh_can_tnm_n

Definition: Describes the absence or presence and extent of regional lymph node metastasis, as classified by TNM, and is specific to the site, TNM is an international staging system comprising:

- pathological staging based on histology reports, and
- clinical staging based on hospital events

Format:

Name of classification:

Notes: Refer to TNM Classification of Malignant Tumours for further details. Introduced in 2001.

IDI variable name: moh_can_tnm_t

Definition: An international classification code to describe the extent/size of the primary tumour, as classified by TNM,tastases, and is specific to the site. TNM is an international staging system comprising:

- pathological staging based on histology reports, and
- clinical staging based on hospital events

Format:

Name of classification:

Notes: Refer to TNM Classification of Malignant Tumours for further details. Introduced in 2001.
