Changes to the NSW cancer incidence and mortality dataset for 2016 release

Date updated: 9th March 2020

The 2016 cancer incidence and mortality data release incorporates a change in the software and technology used to generate our cancer incidence and mortality reporting datasets. Our dataset production has moved from a stand-alone SAP Data Services process, to a centralised enterprise data warehouse solution, the Cancer Institute NSW Data Warehouse.

This change has subsequently led to changes in: geocoding software and methodology; the availability of geographical boundaries; the reference data reported against data elements; and the naming and format of data elements. The changes are described in detail below.

1. Geocoding software and methodology

1.1. Change description

Prior to the release of the 2016 cancer incidence and mortality data, the generation of geocodes (codes for a geographic location) and assignment of geographical boundaries to case address at diagnosis, were produced using software and methodology developed by the NSW Ministry of Health (MoH).

For the 2016 cancer incidence and mortality data release, the principal geocoding process has been converted to an Intech software solution, customised to Cancer Institute NSW requirements. The new solution was applied to the entire database, in order to report a cohesive set of geographical boundaries for cancer cases diagnosed between 1972 and 2016.

Where possible the new process and methodology has been aligned to the previous one, so for the majority of cases there has been minimal impact to typical high-level reporting geographical areas such as local health district (LHD), primary health network (PHN) and local government areas (LGA). There is a higher degree of change impacting lower level geographical areas such as mesh block, Statistical Area 1 (SA1) and Statistical Area 2 (SA2). A summary of the changes impacting geographies available for research and reporting from the different geocoding systems can be found below in section '1.3 – change analysis'.

1.2. Impacted data elements

The following NSW cancer incidence and mortality dataset elements were impacted by the change in geocoding process:

- Local health district
- Primary health network
- LGA 2006 (*Australian Standard Geographical Classification (ASGC))*
- SLA 2006 *(ASGC)*
- LGA 2016 *(Australian Statistical Geography Standard (ASGS))*
- SA2 2016 *(ASGS)*

Record Number: E20/04726

Level 9, 8 Central Avenue, Australian Technology Park, Eveleigh, NSW 2015. PO Box 41, Alexandria, NSW 1435
T 02 8374 5600 • F 02 8374 3600 • www.cancerinstitute.org.au • ABN 48 538 442 594

- SA3 2016 *(ASGS)*
- SA4 2016 *(ASGS)*
- GCCSA 2016 *(ASGS)*
- Postcode
- Remoteness *(ASGC, ASGS)*
- Socioeconomic position - IRSAD deciles *(ASGC, ASGS)*
- Socioeconomic position - IRSAD quintiles *(ASGC, ASGS)*
- Socioeconomic position - IRSD deciles *(ASGC, ASGS)*
- Socioeconomic position - IRSD quintiles *(ASGC, ASGS)*

1.3. Change analysis

The following table shows the proportion of cases for the diagnosis period 1972 to 2016 with a difference in value due to the change in geocoding system described above.

Data element	Proportion of cases with difference (%)
Local health district	0.2
Primary health network	0.2
LGA 2006 *(ASGC)*	0.8
SLA 2006 *(ASGC)*	1
LGA 2016 *(ASGS)*	0.6
SA2 2016 *(ASGS)*	2.1
SA3 2016 *(ASGS)*	0.4
SA4 2016 *(ASGS)*	0.2
GCCSA 2016 *(ASGS)*	0.1
Postcode	1.2
Remoteness *(ASGC, ASGS)*	0.2
Socioeconomic position – IRSAD deciles *(ASGC, ASGS)*	3.4
Socioeconomic position – IRSAD quintiles *(ASGC, ASGS)*	2.5
Socioeconomic position – IRSD deciles *(ASGC, ASGS)*	6.9
Socioeconomic position – IRSD quintiles *(ASGC, ASGS)*	5.5
NSW total row count	**0**

Note: the changes for socioeconomic position are more pronounced, as the calculation is based on a combination of lower level geographical areas (ASGC Collection District (CD) and ASGS Statistical Area 1 (SA1)), and the associated indexes of socioeconomic position.

Level 9, 8 Central Avenue, Australian Technology Park, Eveleigh, NSW 2015. PO Box 41, Alexandria, NSW 1435
T 02 8374 5600 • F 02 8374 3600 • www.cancerinstitute.org.au • ABN 48 538 442 594

A further breakdown of local health district is shown below.

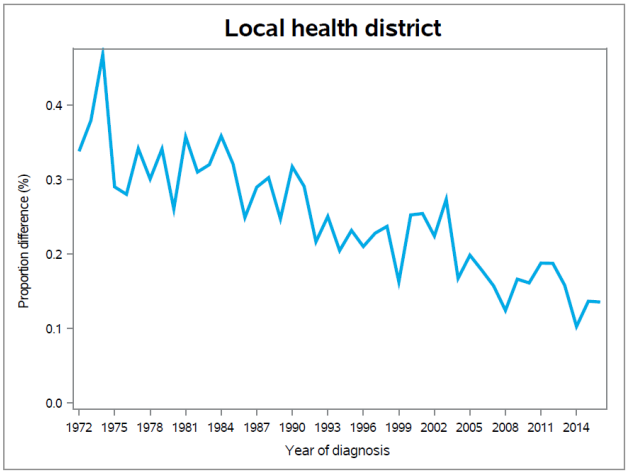
Data element: Local health district	Proportion of cases with difference (%)
Sydney LHD	0.3
South Western Sydney LHD	0.2
South Eastern Sydney LHD	0.1
Illawarra Shoalhaven LHD	0.1
Western Sydney LHD	0.5
Nepean Blue Mountains LHD	0.3
Northern Sydney LHD	0.3
Central Coast LHD	0.1
Hunter New England LHD	0.1
Northern NSW LHD	0.1
Mid North Coast LHD	0.1
Southern NSW LHD	0.2
Murrumbidgee LHD	0.3
Western NSW LHD	0.3
Far West LHD	0.4
NSW Unknown LHD	7.6
Albury residents	0.2

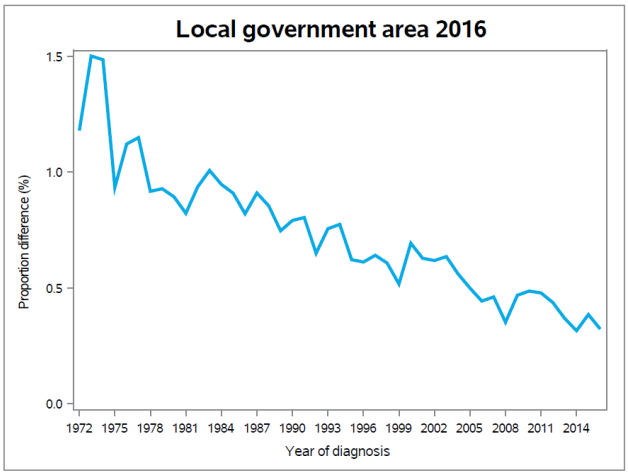
Note: the changes for NSW Unknown LHD are more pronounced due to the small volumes in this category.

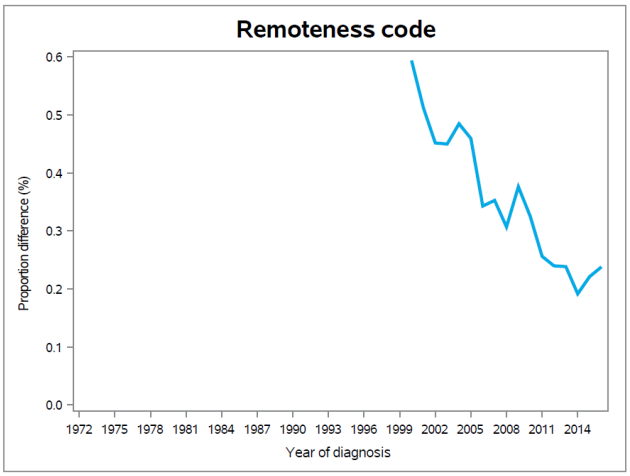
Level 9, 8 Central Avenue, Australian Technology Park, Eveleigh, NSW 2015. PO Box 41, Alexandria, NSW 1435
T 02 8374 5600 • F 02 8374 3600 • www.cancerinstitute.org.au • ABN 48 538 442 594

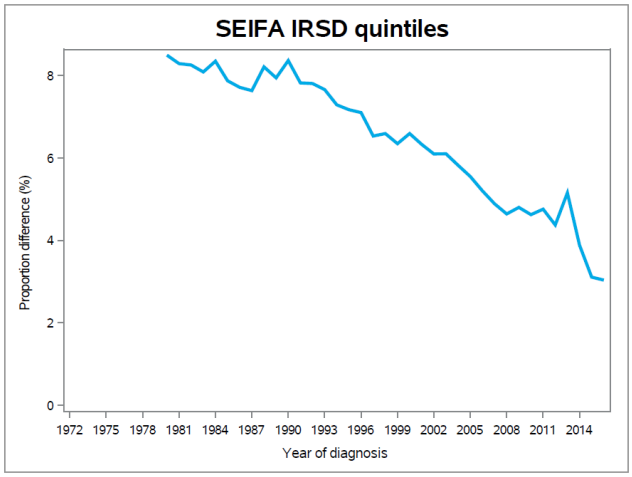
The following charts show the proportion of cases impacted by the change in geocoding methodology trend from 1972 to 2016, for the most commonly used geographical data elements. Historically, there is a greater degree of change and impact. This can be further explained by the less accurate address matching that can occur for older addresses during the address matching phase of the geocoding process. During the address matching phase, addresses are matched to a database of 'current' addresses. We have found that older addresses may not align well to 'current' addresses, as they may have changed in part since the original case address was captured (i.e. changes to localities and/or postcodes). This may lead to the assignment of different geocodes and geographical boundaries.

Other geographical data elements not shown follow a similar pattern of decreasing change over time.









Record Number: E20/04726

Level 9, 8 Central Avenue, Australian Technology Park, Eveleigh, NSW 2015. PO Box 41, Alexandria, NSW 1435
T 02 8374 5600 • F 02 8374 3600 • www.cancerinstitute.org.au • ABN 48 538 442 594

2. Geographical boundaries superseded

2.1. Socioeconomic position – IRSD quintiles (ASGC)

2.1.1. Change description

The "Socioeconomic position – IRSD quintiles (ASGC)" data element is no longer available. It was based entirely on the outdated historical ABS geography classification of ASGC.

It has been superseded by "Socioeconomic position – IRSD quintiles (ASGC, ASGS)". This element has been updated to reflect the current ABS statistical geography classification of ASGS.

For further details about calculation methodology please see the NSW Cancer Registry data dictionary.

2.2. Remoteness 2006 (ASGC)

2.2.1. Change description

The "Remoteness 2006 (ASGC)" data element is no longer available. It was based entirely on the outdated historical ABS geography classification of ASGC.

It has been superseded by "Remoteness (ASGC, ASGS)". This element has been updated to reflect the current ABS statistical geography classification of ASGS.

For further details about calculation methodology please see the NSW Cancer Registry data dictionary.

Level 9, 8 Central Avenue, Australian Technology Park, Eveleigh, NSW 2015. PO Box 41, Alexandria, NSW 1435
T 02 8374 5600 • F 02 8374 3600 • www.cancerinstitute.org.au • ABN 48 538 442 594

3. Reference data

3.1. Date of birth unknown values

3.1.1. Change description

The unknown value for date of birth has changed. This impacts the following data elements:
- Year of birth
- Month of birth
- Day of birth

Please note the reference data value changes and mapping below:

Old data reference values	New data reference values
17 Nov 1858	01 Jan 9999

3.2. Date of diagnosis unknown values

3.2.1. Change description

The unknown value for date of diagnosis has changed. This impacts the following data elements:
- Year of diagnosis
- Month of diagnosis
- Day of diagnosis

Please note the reference data value changes and mapping below:

Old data reference values	New data reference values
11 Nov 1888	01 Jan 9999

Level 9, 8 Central Avenue, Australian Technology Park, Eveleigh, NSW 2015. PO Box 41, Alexandria, NSW 1435
T 02 8374 5600 • F 02 8374 3600 • www.cancerinstitute.org.au • ABN 48 538 442 594

3.3. General missing/unknown values

3.3.1. Change description

The general missing/unknown value for data elements has changed. For impacted data elements please see the NSW Cancer Registry data dictionary.

Please note the reference data changes and mapping below:

Old data reference values	New data reference values
Blank/NULL	-1

3.4. Degree of spread at diagnosis

3.4.1. Change description

The "Degree of spread at diagnosis" data element now has "Regional spread, adjacent organs and/or regional lymph nodes" broken down into "Regional spread, adjacent organs" and "Regional spread, regional lymph nodes" to allow for more detailed analyses.

Please note the reference data value changes and mapping below:

Old data reference values	New data reference values
1 = Localised to tissue of origin	1 = Localised to tissue of origin
2 = Regional spread, adjacent organs and/or regional lymph nodes	2 = Regional spread, adjacent organs
	3 = Regional spread, regional lymph nodes
3 = Distant metastases	4 = Distant metastases
6 = In-situ	6 = In-situ
9 = Unknown	9 = Unknown

Level 9, 8 Central Avenue, Australian Technology Park, Eveleigh, NSW 2015. PO Box 41, Alexandria, NSW 1435
T 02 8374 5600 • F 02 8374 3600 • www.cancerinstitute.org.au • ABN 48 538 442 594

3.5. Best basis of diagnosis

3.5.1. Change description

The "Best basis of diagnosis" reference data have changed.

Please note the reference data value changes and mapping below:

Old data reference values	New data reference values
0 = Cytology including FNA, smears, washing, sputum	5 = Cytology
1 = Clinical/imaging/ biochemical	2 = Clinical
2 = Histopathology performed	7 = Histology performed
5 = Death certificate only	0 = Death certificate only
6 = Histopathology sighted at NSWCR	66 = Histopathology sighted at NSWCR
	9 = Unknown

Note: there are no data with "Unknown" (9) value.

Level 9, 8 Central Avenue, Australian Technology Park, Eveleigh, NSW 2015. PO Box 41, Alexandria, NSW 1435
T 02 8374 5600 • F 02 8374 3600 • www.cancerinstitute.org.au • ABN 48 538 442 594

4. Naming and format of data elements

4.1. Change description

The naming of data elements and some data types and formats have changed as shown in the mapping table below:

Item No.	NSWCR Data Dictionary name (Element)	Old (SAP Data Services)				New (Cancer Institute NSW Data Warehouse)			
		Variable name	Type	Length	Format	Variable name	Type	Length	Format
Demographic data elements									
1	Gender	P_SEX	Char	1	$1	SexName	Char	2	$2
2	Country of birth	P_SACC	Char	4	$4	CountryBirthCode	Char	4	$4
3	Aboriginal and Torres Strait Islander status	P_ABTSI	Char	1	$1	AboriginalityCode	Num	8	2
4	Year of birth	P_YB	Char	4	$4	BirthDateYear	Char	4	$4
5	Month of birth	P_MB	Char	2	$2	BirthDateMonth	Char	2	$2
6	Day of birth	P_DB	Char	2	$2	BirthDateDay	Char	2	$2
7	Date of birth validity code	p_dob_valcode	Num	8	8	BirthDateValidityCode	Num	8	2
Cancer diagnosis data elements									
8	Year of diagnosis	C_YDG	Char	4	$4	DiagnosisDateYear	Char	4	$4
9	Month of diagnosis	C_MDG	Char	2	$2	DiagnosisDateMonth	Char	2	$2
10	Day of diagnosis	C_DDG	Char	2	$2	DiagnosisDateDay	Char	2	$2
11	Date of diagnosis validity code	C_DOD_VALCODE	Char	5	$5	DiagnosisDateValidityCode	Num	8	2
12	Age at diagnosis	C_AGE	Num	8	8	DiagnosisAge	Num	8	3
13	Cancer type	C_TOPOTAB	Char	5	$5	DiagnosisICD9GrpTopoCode	Char	5	$5
14	Clinical cancer group	C_CLINGRP	Num	8	8	DiagnosisClinicalCancerGroupCode	Num	8	2

Level 9, 8 Central Avenue, Australian Technology Park, Eveleigh, NSW 2015. PO Box 41, Alexandria, NSW 1435
T 02 8374 5600 • F 02 8374 3600 • www.cancerinstitute.org.au • ABN 48 538 442 594

Item No.	NSWCR Data Dictionary name (Element)	Old (SAP Data Services)				New (Cancer Institute NSW Data Warehouse)			
		Variable name	Type	Length	Format	Variable name	Type	Length	Format
15	Topography code (ICD-O-3)	C_TOPO4	Char	4	$4	DiagnosisICDO3TopoCode4	Char	4	$4
16	Topography code (ICD-10-AM)	C_TOPO10	Char	4	$4	DiagnosisICD10RepTopoCode	Char	4	$4
17	Morphology code (ICD-O-3)	C_HIST3	Char	3	$3	DiagnosisICDO3MorphCode3	Char	3	$3
18	Morphology code 4 digit (ICD-O-3)	C_HIST4	Char	4	$4	DiagnosisICDO3MorphCode4	Char	4	$4
19	Behaviour code	C_BEHAVE	Char	1	$1	BehaviourCode	Num	8	2
20	Best basis of diagnosis	C_METHOD	Char	1	$1	DiagnosisBasisCode	Num	8	2
21	Degree of spread at diagnosis	C_STAGE	Char	1	$1	ExtentGroup2Code	Num	8	2
22	Laterality	C_LATERALITY	Char	1	$1	LateralityName	Char	3	$3
23	Breslow thickness of melanoma / Size of breast cancer	C_THICKNESS	Num	8	11.3	ThicknessSize	Num	8	7.3
24	Number of primary sites	P_NPRI	Num	8	8	PrimarySiteCount	Num	8	2
25	Registry derived-stage (STaR)	c_Rdstage	Num	8	8	DiagnosisCRDerivedStageCode	Num	8	2
		c_RDstage_desc	Char	30	$30	DiagnosisCRDerivedStageName	Char	30	$30
26	Registry derived staging basis (STaR)	c_Rdstagebasis	Char	1	$1	DiagnosisCRDerivedBasisCode	Char	1	$1
		c_RDstagebasis_desc	Char	20	$20	DiagnosisCRDerivedBasisName	Char	20	$20
Mortality data elements									
27	Year of death	P_YDTH	Char	4	$4	DeathDateYear	Char	4	$4
28	Month of death	P_MDTH	Char	2	$2	DeathDateMonth	Char	2	$2
29	Day of death	P_DDTH	Char	2	$2	DeathDateDay	Char	2	$2
30	Age at death	P_AGED	Num	8	8	DeathAge	Num	8	3
31	Cause of death cancer type	P_CAUSETAB2	Char	5	$5	DeathICD9GrpTopoCode	Char	5	$5

Record Number: E20/04726

Level 9, 8 Central Avenue, Australian Technology Park, Eveleigh, NSW 2015. PO Box 41, Alexandria, NSW 1435
T 02 8374 5600 • F 02 8374 3600 • www.cancerinstitute.org.au • ABN 48 538 442 594

Item No.	NSWCR Data Dictionary name (Element)	Old (SAP Data Services)				New (Cancer Institute NSW Data Warehouse)			
		Variable name	Type	Length	Format	Variable name	Type	Length	Format
32	Cause of death clinical cancer group	P_CLINMGRP2	Num	8	8	DeathClinicalCancerGroupCode	Num	8	2
33	Cause of death topography code (ICD-O-3)	P_CAUSE4	Char	4	$4	DeathICDO3TopoCode4	Char	4	$4
34	Cause of death topography code (ICD-10-AM)	P_CAUSE10	Char	5	$5	DeathICD10RepTopoCode	Char	5	$5
35	Place of death group	p_place_group	Char	2	$2	PlaceDeathGroupCode	Char	2	$2
Geographical data elements (based on residence at diagnosis)									
36	Postcode	C_GEOPC	Char	4	$4	GCDiagnosisPostcode	Char	4	$4
37	LGA 2006 (ASGC)	C_LGA2006	Num	8	8	GCDiagnosisLGACode2006	Num	8	5
38	SLA 2006 (ASGC)	C_SLA2006	Char	5	$5	GCDiagnosisSLACode2006	Num	8	5
N/A	Remoteness 2006 (ASGC)	C_ARIA_CODE	Char	1		N/A	N/A	N/A	N/A
N/A	Socioeconomic position – IRSD quintiles (ASGC)	C_SEIFA_CODE	Char	1		N/A	N/A	N/A	N/A
39	LGA 2016 (ASGS)	C_LGA2016	Num	8	5	GCDiagnosisLGACode2016	Num	8	5
		C_LGA2016_NAME	Char	50	$50	GCDiagnosisLGAName2016	Char	50	$50
40	SA2 2016 (ASGS)	C_SA22016	Num	8	9	GCDiagnosisSA2Code2016	Num	8	9
		C_SA22016_NAME	Char	50	$50	GCDiagnosisSA2Name2016	Char	50	$50
41	SA3 2016 (ASGS)	C_SA32016	Num	8	5	GCDiagnosisSA3Code2016	Num	8	5
		C_SA32016_NAME	Char	50	$50	GCDiagnosisSA3Name2016	Char	50	$50
42	SA4 2016 (ASGS)	C_SA42016	Num	8	3	GCDiagnosisSA4Code2016	Num	8	3
		C_SA42016_NAME	Char	50	$50	GCDiagnosisSA4Name2016	Char	50	$50

Level 9, 8 Central Avenue, Australian Technology Park, Eveleigh, NSW 2015. PO Box 41, Alexandria, NSW 1435
T 02 8374 5600 • F 02 8374 3600 • www.cancerinstitute.org.au • ABN 48 538 442 594

Item No.	NSWCR Data Dictionary name (Element)	Old (SAP Data Services)				New (Cancer Institute NSW Data Warehouse)			
		Variable name	Type	Length	Format	Variable name	Type	Length	Format
43	GCCSA 2016 (ASGS)	C_GCCSA2016_NAME	Char	50	$50	GCDiagnosisGCCSAName2016	Char	255	$255
44	Remoteness (ASGC, ASGS)	RemotenessCode	Num	8	8	RemotenessCode	Num	8	2
45	Remoteness calculation method	RemotenessCalcMethod	Char	50	$50	RemotenessCalcMethod	Char	50	$50
46	Socioeconomic position – IRSAD deciles (ASGC, ASGS)	SEIFAIRSADDecileCode	Num	8	best12	SEIFAIRSADDecileCode	Num	8	2
47	Socioeconomic position – IRSAD quintiles (ASGC, ASGS)	SEIFAIRSADQuintileCode	Num	8	best12	SEIFAIRSADQuintileCode	Num	8	2
48	Socioeconomic position – IRSD deciles (ASGC, ASGS)	SEIFAIRSDDecileCode	Num	8	best12	SEIFAIRSDDecileCode	Num	8	2
49	Socioeconomic position – IRSD quintiles (ASGC, ASGS)	SEIFAIRSDQuintileCode	Num	8	best12	SEIFAIRSDQuintileCode	Num	8	2
50	Socioeconomic position calculation method	SEIFACalcMethod	Char	50	$50	SEIFACalcMethod	Char	50	$50
51	Local health district	C_LHD2011	Char	4	$4	GCDiagnosisLHDCode2010	Char	4	$4
		lhd_desc	Char	25	$25	GCDiagnosisLHDDescription2010	Char	25	$25
52	Primary health network	N/A	N/A	N/A	N/A	GCDiagnosisPHNCode2015	Char	6	$6
		phn2015_desc	Char	40	$40	GCDiagnosisPHNDescription2015	Char	40	$40

Level 9, 8 Central Avenue, Australian Technology Park, Eveleigh, NSW 2015. PO Box 41, Alexandria, NSW 1435
T 02 8374 5600 • F 02 8374 3600 • www.cancerinstitute.org.au • ABN 48 538 442 594