NSW Cancer Registry – Data dictionary - NSW

Date updated: 09 November 2021

Data set: NSW Cancer Registry (NSWCR) 2018 NSW cancer incidence and mortality data.

Data available:

- Invasive primary cancer cases:
    - Incidence: 1972-2018
    - Mortality: 1972-2018
- In-situ cancer cases for breast cancer and melanoma only:
    - Incidence: 2002-2018
    - Mortality: 2002-2018
- Clinical episodes of care: 2013 only

Item no.	Data element	Description/notes	Data values/format
Demographic data elements			
1.	Gender	The gender of the person.	M = Male F = Female
2.	Country of birth	The country in which the person was born as represented by a code. The Standard Australian Classification of Countries (SACC 2nd edition), issued by the Australian Bureau of Statistics. Caveat: Country of birth is incomplete on the NSWCR. Country of birth is updated from hospital inpatient records if it is available. Particular cancer sites like melanoma, where a person does not always have a hospital visit and the pathology report is the main source of notification, will have a larger proportion of unknown country of birth. For the most recent 5-year period of data available (2014-2018), approximately 14% of diagnoses had unknown country of birth. Country of birth for persons **diagnosed in 2014** were impacted by a system issue, producing an increased proportion of **unknown** country of birth.	http://www.abs.gov.au/AUSSTATS/abs@.nsf/Latestproducts/2547A73B811A360DCA25744B00153861?opendocument 0000 = Unknown COB code -1 = Not available (invalid COB code)

Item no.	Data element	Description/notes	Data values/format
3.	Aboriginality	The Aboriginality code for the person. Permission is required from the **Aboriginal Health and Medical Research Council** when publishing reports utilising this data element. Caveat: Aboriginality is incomplete on the NSW Cancer Registry (NSWCR). This data element is updated from hospital inpatient records if available and ABS coded cause of death data. This affects primarily incidence calculations. From 1999 to 2018, approximately 10% of all NSW invasive cancer cases recorded on the NSWCR, did not have Aboriginality recorded (unknown or not available). Aboriginality for persons **diagnosed in 2014** were impacted by a system issue, producing an increased proportion of **unknown** Aboriginality. Periods of use: 1999 to latest available (for incidence) 1994 to latest available (for mortality)	1 = Aboriginal 2 = Torres Strait Islander 3 = Both Aboriginal and Torres Strait Islander 4 = Neither 8 = Declined 9 = Unknown -1 = Not available
4.	Year of birth	The year of birth of the person.	Format: YYYY *Unknown* year of birth = 9999
5.	Month of birth	The month of birth of the person.	Format: MM *Unknown* month of birth = Jan 9999
6.	Day of birth	The day of birth of the person.	Format: DD *Unknown* date of birth = 01 Jan 9999
7.	Date of birth validity code	A code reflecting the accuracy of a person's date of birth. Date of birth was not a mandatory item prior to 1994.	1 = Actual date (day, month, year known) 2 = Date unknown 3 = Day estimated (month, year known) 4 = Day and month estimated (year known) -1 = Not available

Item no.	Data element	Description/notes	Data values/format
Cancer diagnosis data elements			
8.	Year of diagnosis	The year on which the person was first diagnosed with cancer. Date of diagnosis reflects the earliest definitive diagnostic or treatment episode received by the NSW Cancer Registry and may not be the exact date of diagnosis. <u>Caveat</u> Please refer to caveat document for further details.	Format: YYYY *Unknown* year of diagnosis = 9999
9.	Month of diagnosis	The month on which the person was first diagnosed with cancer. Date of diagnosis reflects the earliest definitive diagnostic or treatment episode received by the NSW Cancer Registry and may not be the exact date of diagnosis. <u>Caveat</u> Please refer to caveat document for further details.	Format: MM *Unknown* month of diagnosis = Jan 9999
10.	Day of diagnosis	The day on which the person was first diagnosed with cancer. Date of diagnosis reflects the earliest definitive diagnostic or treatment episode received by the NSWCR and may not be the exact date of diagnosis. <u>Caveat</u> Please refer to caveat document for further details.	Format: DD *Unknown* date of diagnosis = 01 Jan 9999
11.	Date of diagnosis validity code	A code reflecting the accuracy of a date of diagnosis.	1 = Actual date (day, month, year known) 2 = Date unknown 3 = Day estimated (month, year known) 4 = Day and month estimated (year known) -1 = Not available
12.	Age at diagnosis	The age at which the person was first diagnosed with cancer. Calculated from date of birth and date of diagnosis.	-1 = Not available (unknown age at diagnosis) Format: NNN

Item no.	Data element	Description/notes	Data values/format
13.	Cancer type	The type of cancer according to the reporting categories for cancer types. The tabulation category is based on ICD-9 and is derived from ICD-O-3 topography and morphology.	See Attachment 1 – Cancer type reporting categories Format: CNNNN
14.	Clinical cancer group	The clinical cancer group of the cancer at diagnosis. Clinical cancer groups are a grouping of cancers based on treatment categories developed with advice from Professor Jim Bishop and endorsed by the Australasian Association of Cancer Registries (AACR) in December 2004. Clinical group is derived from the grouping of cancer types defined in element 13.	See Attachment 2 – Clinical cancer group Format: NN
15.	Topography code (ICD-O-3)	The anatomical site of origin (or organ system) of the cancer at diagnosis. Coded according to the WHO International Classification of Diseases for Oncology (third edition).	https://www.who.int/standards/classifications/other-classifications/international-classification-of-diseases-for-oncology Format: CNNN
16.	Topography code (ICD-10-AM)	The anatomical site of origin (or organ system) of the cancer at diagnosis. Coded according to the International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, Australian Modification (ICD-10-AM). Australian modification of the WHO ICD-10 base classification system, modified to serve particular Australian needs and to support the national collection of data relevant to the population's health.	https://www.ihacpa.gov.au/resources/icd-10-amachiacs-twelfth-edition -1 = Not available (unknown topography code) Format: CNNN

Item no.	Data element	Description/notes	Data values/format
17.	Morphology code (ICD-O-3)	The morphology code of the cancer at diagnosis. Morphology codes describe the cell type of the cancer according to the WHO International Classification of Diseases for Oncology (third edition). A three-digit morphology code refers to the main grouping by cell type e.g. Lymphoid leukaemia. Caveat Morphology codes may change over time as classifications change. E.g. Myelodysplasia was previously considered to have uncertain tumour behaviour, but since 2001 this element has been considered an invasive cancer. Data element is available from 2003 year of diagnosis.	https://www.who.int/standards/classifications/other-classifications/international-classification-of-diseases-for-oncology Format: NNN
18.	Morphology code 4 digit (ICD-O-3)	The morphology code of the cancer at diagnosis. Morphology codes describe the cell type of the cancer according to the WHO International Classification of Diseases for Oncology (third edition). A four-digit morphology code contains a more specific morphology e.g. Acute lymphoblastic leukaemia, not otherwise specified. Caveat Morphology codes may change over time as classifications change. E.g. Myelodysplasia was previously considered to have uncertain tumour behaviour, but since 2001 this element has been considered an invasive cancer. Data element is available from 2003 year of diagnosis.	https://www.who.int/standards/classifications/other-classifications/international-classification-of-diseases-for-oncology Format: NNNN
19.	Behaviour code	The behaviour of the cancer at diagnosis. It is equivalent to the 5th digit of the ICD-O-3 morphology code. Available for all invasive cancers and in-situ cases for breast and melanoma only. In-situ cases are complete and available from 2002 onwards.	2 = In-situ 3 = Invasive
20.	Best basis of diagnosis	Best basis of diagnosis or method of diagnosis is a summarised field based on the highest level of verification of cancer specimens within a four-month period.	0 = Death certificate only 2 = Clinical 5 = Cytology 7 = Histology performed 9 = Unknown 66 = Histopathology sighted at NSWCR

Item no.	Data element	Description/notes	Data values/format
		For example, if a case is originally notified through cytology and a pathology report is received and read by a NSW Cancer Registry (NSWCR) coder, then the method at diagnosis is 'Histopathology sighted at NSWCR'.	
21.	Degree of spread at diagnosis	Degree of spread at diagnosis assigned from the highest degree of spread (extent of spread of cancer from its point of origin) within four months of the cancer diagnosis. Caveat Please refer to caveat document for further details.	1 = Localised to tissue of origin 2 = Regional spread, adjacent organs 3 = Regional spread, regional lymph nodes 4 = Distant metastases 6 = In-situ 9 = Unknown
22.	Laterality	Laterality describes which side of a paired organ is the origin of the primary cancer. Each side of a paired organ is considered separately and described as lateral when occurring, unless a physician determines that it is bilateral. A paired organ is one in which there are two separate organs of the same kind, one on either side of the body (e.g. kidney, breast, ovary, testis and lung).	L = Left R = Right UNK = Unknown (Not applicable or not available)
23.	Breslow thickness of melanoma / Size of breast cancer	**Cutaneous melanoma thickness** is recorded completely for all cases diagnosed in 1988 onwards. Thickness (mm) is the thickness of the total tissue extracted. For **breast cancer**, tumour size is recorded completely for all cases diagnosed in 1997 only and from 2004 onwards. Size (mm) is the size of the total breast cancer tissue extracted.	Valid values are 0.1-994 (mm). Other values are: 995 = Micro-invasion 996 = Size < 0.1 mm 997 = Size described as small 998 = Size described as big 999 = Unknown 888 = Size not recorded due to neoadjuvant therapy (breast cancer only) -1 = Missing/unavailable Format NNN.NNN
24.	Number of primary sites	Primary sites are defined as the number of invasive primary cancer cases reported for statistical purposes in the NSW reporting data set. Cancer cases are subjected to the international rules for determining multiple primary cancers developed by the International Association of Cancer Registries.	Format: NN

Item no.	Data element	Description/notes	Data values/format
25.	Registry derived stage (STaR)	Simplified TNM stage at diagnosis derived from data sources available to the NSW Cancer Registry and collected as part of Cancer Australia's Stage, Treatment and Recurrence (STaR) project. Cancer Australia has supported the Cancer Council Victoria to develop business rules to derive TNM stage at diagnosis based on data sources available to all population-based cancer registries. The rules were endorsed by the Australasian Association of Cancer Registries (AACR). Registry derived stage is available for **2011 year of diagnosis** only, for the five highest incidence cancers (**prostate, breast, lung, bowel and melanoma**).	1 = Stage I 2 = Stage II 3 = Stage III 4 = Stage IV 88 = Stage not applicable 9 = Stage missing/unavailable (includes morphologies not eligible for TNM staging and tumours with no histological confirmation (as per AJCC rules)) -1 = Not available (year of diagnosis not 2011) Cancer Australia: https://canceraustralia.gov.au/research-data/cancer-data/improving-cancer-data
26.	Registry derived staging basis (STaR)	Accompanies registry derived stage to clarify the source of the information used in the derivation of stage for a particular tumour.	C = Clinical only I = Integrated (combined clinical and pathological stage) P = Pathology only U = Unknown -1 = Not available (year of diagnosis not 2011)
Mortality data elements			
27.	Year of death	The year of death of the person.	-1 = Not available (person alive or unknown date of death) Format: YYYY
28.	Month of death	The month of death of the person.	-1 = Not available (person alive or unknown date of death) Format: MM
29.	Day of death	The day of death of the person.	-1 = Not available (person alive or unknown date of death) Format: DD
30.	Age at death	The age of the person at death.	-1 = Not available (person alive or unknown age at death) Format: NNN

Item no.	Data element	Description/notes	Data values/format
31.	Cause of death cancer type	The cancer type that has led to the death of the person, according to the reporting categories for cancer types. The tabulation category is based on ICD-9 and is derived from the ICD-O-3 topography and morphology that has caused the death.	See Attachment *1* – Cancer type reporting categories -1 = Not available (person alive or unknown cancer type) Format: CNNNN
32.	Cause of death clinical cancer group	The clinical cancer group that has led to the death of the person. Clinical cancer groups are a grouping of cancers based on treatment categories developed with advice from Professor Jim Bishop and endorsed by the Australasian Association of Cancer Registries (AACR) in December 2004. Cause of death clinical cancer group is derived from the grouping of cause of death cancer types in Item 31.	See Attachment 2 – Clinical cancer group -1 = Not available (person alive or unknown clinical cancer group) Format: NN
33.	Cause of death topography code (ICD-O-3)	The anatomical site of origin (or organ system) of the cancer that has caused the death of the person. Coded according to the WHO International Classification of Diseases for Oncology (third edition).	https://www.who.int/standards/classifications/other-classifications/international-classification-of-diseases-for-oncology -1 = Not available (person alive or unknown topography code) Format: CNNN
34.	Cause of death topography code (ICD-10-AM)	The anatomical site of origin (or organ system) of the cancer that has caused the death of the person. Coded according to the International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, Australian Modification (ICD-10-AM). Australian modification of the WHO ICD-10 base classification system, modified to serve particular Australian needs and to support the national collection of data relevant to the population's health.	https://www.ihacpa.gov.au/resources/icd-10-amachiacs-twelfth-edition -1 = Not available (person alive or unknown topography code) Format: CNNN
35.	Place of death group	Persons that have been diagnosed with cancer in NSW and have subsequently died of cancer in NSW have been grouped into specific categories of place of death as approved by the NSW Registry of Births Deaths and Marriages.	1 = Died at Home 2 = Died in a Hospice 3 = Died in a Public Hospital 4 = Died in a Private Hospital 5 = Died in a Nursing Home

Item no.	Data element	Description/notes	Data values/format
		Caveat Please refer to caveat document for further details.	9 = Unknown Place of Death NA = Not applicable (e.g. non-cancer or interstate death) -1 = Not available (person alive or place of death group not available)
Geographical data elements (based on residence at diagnosis)			
36.	Postcode	Postcode of residence at diagnosis assigned by geocoding.	9998 = No fixed address in NSW 9999 = Unknown address in NSW Format: NNNN
37.	LGA 2006 *(ASGC)*	Local government area (LGA) of residence at diagnosis. Sourced from geocoding. Based on the Australian Standard Geographical Classification (ASGC 2006), issued by the Australian Bureau of Statistics (ABS), effective at 1st July 2006. The LGA is a spatial unit which represents the whole geographical area of responsibility of an incorporated, officially gazetted local government area/council, as defined by each state and territory.	http://www.abs.gov.au/AUSSTATS/abs@.nsf/Previousproducts/943E92C473A3C463CA2571A900177C33?opendocument -1 = Not available (unknown LGA code) Format: NNNNN
38.	SLA 2006 *(ASGC)*	Statistical local area (SLA) of residence at diagnosis. Sourced from geocoding (rolled up from collection district (CD)). Based on the Australian Standard Geographical Classification (ASGC 2006), issued by the Australian Bureau of Statistics (ABS), effective at 1st July 2006. The SLA is a base spatial unit used to collect and disseminate statistics other than those collected from the population censuses. In non-census years, the SLA is the smallest unit defined in the ASGC. In census years, an SLA consists of one or more whole CDs.	https://www.abs.gov.au/AUSSTATS/abs@.nsf/Previousproducts/4AD66A34D23E5EB0CA2571A9001776AB?opendocument -1 = Not available (unknown SLA code) Format: NNNNN
39.	LGA 2016 *(ASGS)*	Local government area (LGA) of residence at diagnosis. Sourced from geocoding. Based on the Australian Statistical Geography Standard (ASGS 2016), issued by the Australian Bureau of Statistics (ABS), effective at 1st July 2016.	http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/1270.0.55.003~July%202016~Main%20Features~Local%20Government%20Areas%20(LGA)~7 -1 = Not available (unknown LGA code)

Item no.	Data element	Description/notes	Data values/format
		The LGA is a spatial unit which represents the whole geographical area of responsibility of an incorporated, officially gazetted local government area or council, as defined by each state and territory.	Format: NNNNN
40.	SA2 2016 (*ASGS*)	Statistical area level 2 (SA2) of residence at diagnosis. Sourced from geocoding (rolled up from mesh block (MB)). Based on the Australian Statistical Geography Standard (ASGS 2016), issued by the Australian Bureau of Statistics (ABS), effective at 1st July 2016. The SA2 is a general-purpose medium-sized spatial unit built from whole SA1s. Their aim is to represent a community that interacts together socially and economically.	https://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/1270.0.55.001~July%202016~Main%20Features~Statistical%20Area%20Level%202%20(SA2)~10014 -1 = Not available (unknown SA2 code) Format: NNNNNNNNN
41.	SA3 2016 (*ASGS*)	Statistical area level 3 (SA3) of residence at diagnosis. Sourced from geocoding (rolled up from mesh block (MB)). Based on the Australian Statistical Geography Standard (ASGS 2016), issued by the Australian Bureau of Statistics (ABS), effective at 1st July 2016. The SA3 is a regional spatial unit built from whole SA2s. Their aim is to create a standard framework for the analysis of ABS data at the regional level through grouping SA2s with similar regional characteristics.	https://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/1270.0.55.001~July%202016~Main%20Features~Statistical%20Area%20Level%203%20(SA3)~10015 -1 = Not available (unknown SA3 code) Format: NNNNN
42.	SA4 2016 (*ASGS*)	Statistical area level 4 (SA4) of residence at diagnosis. Sourced from geocoding (rolled up from mesh block (MB)). Based on the Australian Statistical Geography Standard (ASGS 2016), issued by the Australian Bureau of Statistics (ABS), effective at 1st July 2016. The SA4 is a standardised spatial unit built from whole SA3s. They are designed for the output of	https://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/1270.0.55.001~July%202016~Main%20Features~Statistical%20Area%20Level%204%20(SA4)~10016 -1 = Not available (unknown SA4 code) Format: NNN

Item no.	Data element	Description/notes	Data values/format
		labour force data and reflect labour markets within each state and territory.	
43.	GCCSA 2016 (*ASGS*)	Greater capital city statistical area (GCCSA) of residence at diagnosis. Sourced from geocoding (rolled up from mesh block (MB)). Based on the Australian Statistical Geography Standard (ASGS 2016), issued by the Australian Bureau of Statistics (ABS), effective at 1st July 2016. The GCCSAs represent the socio-economic extent of state and territory capital cities and are built up from whole SA4s.	https://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/1270.0.55.001~July%202016~Main%20Features~Greater%20Capital%20City%20Statistical%20Areas%20(GCCSA)~10003 Not available = Unknown GCCSA
44.	Remoteness (*ASGC, ASGS*)	Remoteness of residence at diagnosis. Remoteness is based on the Accessibility/Remoteness Index of Australia (ARIA+) methodology developed by the National Key Centre for Social Applications of GIS (GISCA) and utilised by the Australian Bureau of Statistics (ABS). GISCA is now incorporated by the University of Adelaide's Hugo Centre for Migration and Population Research. It is based on a combination of the Australian Standard Geographical Classification (ASGC) and the Australian Statistical Geography Standard (ASGC), depending on the years of diagnosis: Year of diagnosis Methodology used 2000 to 2010 CD (ASGC 2006) 2011 to 2013 SA1 (ASGS 2011) 2014 to 2018 SA1 (ASGS 2016) Remoteness is assigned to records from **year of diagnosis 2000** and onwards.	0 = Major cities 1 = Inner regional 2 = Outer regional 3 = Remote 4 = Very remote -1 = Not available (prior to year of diagnosis 2000) -2 = No score available (valid boundary) -3 = Unknown boundary (unknown address) Remoteness 2016 (ASGS) http://www.abs.gov.au/ausstats/abs@.nsf/mf/1270.0.55.005
45.	Remoteness calculation method	Describes which ABS classification system and which geographical area has been used to calculate remoteness.	-1 = Not available (prior to year of diagnosis 2000) Format: YYYY-AAN
46.	Socioeconomic position - IRSAD deciles (*ASGC, ASGS*)	Socioeconomic position of residence at diagnosis. Socioeconomic position is a general socio-economic index that summarises a range of information	1 = Decile 1 (most disadvantaged) 2 = Decile 2 3 = Decile 3 4 = Decile 4

Item no.	Data element	Description/notes	Data values/format
		about the economic and social conditions of people and households within a geographic area. The index of relative socioeconomic advantage and disadvantage (IRSAD) is developed by the Australian Bureau of Statistics (ABS). It is based on a combination of the Australian Standard Geographical Classification (ASGC) and the Australian Statistical Geography Standard (ASGC), depending on the years of diagnosis: Year of diagnosis Methodology used 1999 to 2003 CD (ASGC 2001) 2004 to 2010 CD (ASGC 2006) 2011 to 2013 SA1 (ASGS 2011) 2014 to 2018 SA1 (ASGS 2016) Socioeconomic position for IRSAD is assigned to records for **year of diagnosis 1999 and onwards**, using the relevant census year index. Caveat The ABS indicate that the census year indexes are not designed for analysis over time and doing so should not be attempted lightly. For more information please refer to the ABS website: http://www.abs.gov.au/websitedbs/censushome.nsf/home/seifahelpansuis?opendocument&navpos=260	5 = Decile 5 6 = Decile 6 7 = Decile 7 8 = Decile 8 9 = Decile 9 10 = Decile 10 (least disadvantaged) -1 = Not available (prior to year of diagnosis 1980) -2 = No score available (valid boundary but no score for year of diagnosis 1980 to 1998) -3 = Unknown boundary (unknown address) Socioeconomic position (2016) http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/2033.0.55.001~2016~Main%20Features~SEIFA%20Basics~5
47.	Socioeconomic position - IRSAD quintiles *(ASGC, ASGS)*	Socioeconomic position of residence at diagnosis. Socioeconomic position is a general socio-economic index that summarises a range of information about the economic and social conditions of people and households within a geographic area. The index of relative socioeconomic advantage and disadvantage (IRSAD), developed by the Australian Bureau of Statistics (ABS). It is based on a combination of the Australian Standard Geographical Classification (ASGC) and the Australian Statistical Geography Standard (ASGC), depending on the years of diagnosis: Year of diagnosis Methodology used 1999 to 2003 CD (ASGC 2001) 2004 to 2010 CD (ASGC 2006) 2011 to 2013 SA1 (ASGS 2011) 2014 to 2018 SA1 (ASGS 2016)	1 = Quintile 1 (most disadvantaged) 2 = Quintile 2 3 = Quintile 3 4 = Quintile 4 5 = Quintile 5 (least disadvantaged) -1 = Not available (prior to year of diagnosis 1980) -2 = No score available (valid boundary but no score for year of diagnosis 1980 to 1998) -3 = Unknown boundary (unknown address) Socioeconomic position (2016) http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/2033.0.55.001~2016~Main%20Features~SEIFA%20Basics~5

Item no.	Data element	Description/notes	Data values/format
		Socioeconomic position for IRSAD is assigned to records for **year of diagnosis 1999** and onwards, using the relevant census year index. Caveat The ABS indicates that the census year indexes are not designed for analysis over time and doing so should not be attempted lightly. For more information please refer to the ABS website: http://www.abs.gov.au/websitedbs/censushome.nsf/home/seifahelpansuis?opendocument&navpos=260	
48.	Socioeconomic position – IRSD deciles *(ASGC, ASGS)*	Socioeconomic position of residence at diagnosis. Socioeconomic position is a general socio-economic index that summarises a range of information about the economic and social conditions of people and households within a geographic area. The index of relative socioeconomic disadvantage (IRSD) is developed by the Australian Bureau of Statistics (ABS). It is based on a combination of the Australian Standard Geographical Classification (ASGC) and the Australian Statistical Geography Standard (ASGC), depending on the years of diagnosis: Year of diagnosis Methodology used 1980 to 1988 CD (ASGC 1986) 1989 to 1993 CD (ASGC 1991) 1994 to 1998 CD (ASGC 1996) 1999 to 2003 CD (ASGC 2001) 2004 to 2010 CD (ASGC 2006) 2011 to 2013 SA1 (ASGS 2011) 2014 to 2018 SA1 (ASGS 2016) Socioeconomic position for IRSD is assigned to records for **year of diagnosis 1980** and onwards, using the relevant census year index. Caveat The ABS indicates that the census year indexes are not designed for analysis over time and doing so should not be attempted lightly. For more information please refer to the ABS website:	1 = Decile 1 (most disadvantaged) 2 = Decile 2 3 = Decile 3 4 = Decile 4 5 = Decile 5 6 = Decile 6 7 = Decile 7 8 = Decile 8 9 = Decile 9 10 = Decile 10 (least disadvantaged) -1 = Not available (prior to year of diagnosis 1980) -2 = No score available (valid boundary) -3 = Unknown boundary (unknown address) Socioeconomic position (2016) http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/2033.0.55.001~2016~Main%20Features~SEIFA%20Basics~5

Item no.	Data element	Description/notes	Data values/format
		http://www.abs.gov.au/websitedbs/censushome.nsf/home/seifahelpansuis?opendocument&navpos=260	
49.	Socioeconomic position – IRSD quintiles *(ASGC, ASGS)*	Socioeconomic position of residence at diagnosis. Socioeconomic position is a general socio-economic index that summarises a range of information about the economic and social conditions of people and households within a geographic area. The index of relative socioeconomic disadvantage (IRSD) is developed by the Australian Bureau of Statistics (ABS). It is based on a combination of the Australian Standard Geographical Classification (ASGC) and the Australian Statistical Geography Standard (ASGC), depending on the years of diagnosis: Year of diagnosis Methodology used 1980 to 1988 CD (ASGC 1986) 1989 to 1993 CD (ASGC 1991) 1994 to 1998 CD (ASGC 1996) 1999 to 2003 CD (ASGC 2001) 2004 to 2010 CD (ASGC 2006) 2011 to 2013 SA1 (ASGS 2011) 2014 to 2018 SA1 (ASGS 2016) Socioeconomic position for IRSD is assigned to records for **year of diagnosis 1980** and onwards, using the relevant census year index. Caveat The ABS indicates that the census year indexes are not designed for analysis over time and doing so should not be attempted lightly. For more information please refer to the ABS website: http://www.abs.gov.au/websitedbs/censushome.nsf/home/seifahelpansuis?opendocument&navpos=260	1 = Quintile 1 (most disadvantaged) 2 = Quintile 2 3 = Quintile 3 4 = Quintile 4 5 = Quintile 5 (least disadvantaged) -1 = Not available (prior to year of diagnosis 1980) -2 = No score available (valid boundary) -3 = Unknown boundary (unknown address) Socioeconomic position (2016) http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/2033.0.55.001~2016~Main%20Features~SEIFA%20Basics~5
50.	Socioeconomic position calculation method	Describes which ABS classification system and which geographical area has been used to calculate socioeconomic position.	-1 = Not available (prior to year of diagnosis 1980) Format: YYYY-AAN
51.	Local health district	Local health district (LHD) of residence at diagnosis. Sourced from geocoding.	http://www.health.nsw.gov.au/lhd/pages/default.aspx

Item no.	Data element	Description/notes	Data values/format
		The local health district is a NSW Health division of the state public services, as at July 2010. Eight local health districts cover the Sydney metropolitan region, and seven cover rural and regional NSW.	X700 = Sydney LHD X710 = South Western Sydney LHD X720 = South Eastern Sydney LHD X730 = Illawarra Shoalhaven LHD X740 = Western Sydney LHD X750 = Nepean Blue Mountains LHD X760 = North Sydney LHD X770 = Central Coast LHD X800 = Hunter New England LHD X810 = Northern NSW LHD X820 = Mid North Coast LHD X830 = Southern NSW LHD X840 = Murrumbidgee LHD X850 = Western NSW LHD X860 = Far West LHD X910 = Unknown NSW LHD X921 = Albury residents -1 = Not available (unknown LHD) Format: ANNN
52.	Primary health network	Primary health network (PHN) of residence at diagnosis. Sourced from geocoding (rolled up from local health district (LHD)). The primary health network as defined by NSW Health, in July 2015. Ten primary health networks cover the Sydney metropolitan region, rural and regional NSW.	http://www.health.gov.au/internet/main/publishing.nsf/Content/PHN-Home PHN101 = Central and Eastern Sydney PHN PHN102 = Northern Sydney PHN PHN103 = Western Sydney PHN PHN104 = Nepean Blue Mountains PHN PHN105 = South Western Sydney PHN PHN106 = South Eastern NSW PHN PHN107 = Western NSW PHN PHN108 = Hunter New England and Central Coast PHN PHN109 = North Coast PHN PHN110 = Murrumbidgee PHN PHN910 = NSW Unknown PHN -1 = Not available (unknown PHN) Format: AAANNN
Episode of care data elements - these data elements are not population-based and have undergone minimal data validation and quality checks. Data is currently available for episodes of care that occurred in 2013 only. Episodes of care occurring from 2014 onwards will be available in 2022.			
53.	Data source type	The source of the information for the episode of care.	I = Inpatient C = Medical Oncology O = Outpatient P = Pathology

Item no.	Data element	Description/notes	Data values/format
			R = Radiotherapy
54.	Episode modality	The modality of the episode of care.	1 = Surgery 2 = Radiotherapy 3 = Terminal Palliation 4 = Admitted/other 5 = Diagnostic 6 = Terminal palliation
55.	Episode start date	The date the episode of care started.	Format: DDMMYYYY
56.	Episode end date	The date the episode of care ended. End dates for medical oncology treatment may not be available for oral and hormonal therapies that may be taken over a period of years.	Format: DDMMYYYY *Unknown year* = 9999
57.	Facility name	The facility where the episode of care occurred.	Episodes of care occurring in facilities not able to be identified are grouped together as the following codes: Interstate facility Private facility Unknown facility
58.	Local health district of Facility	Local health district of facility	X630 = Sydney Children's Hospitals Network X690 = St Vincent's Health Network X700 = Sydney LHD X710 = South Western Sydney LHD X720 = South Eastern Sydney LHD X730 = Illawarra Shoalhaven LHD X740 = Western Sydney LHD X750 = Nepean Blue Mountains LHD X760 = North Sydney LHD X770 = Central Coast LHD X800 = Hunter New England LHD X810 = Northern NSW LHD X820 = Mid North Coast LHD X830 = Southern NSW LHD X840 = Murrumbidgee LHD X850 = Western NSW LHD X860 = Far West LHD UNKN = LHD unknown Episodes of care occurring in facilities not able to be identified are grouped together with the following codes:

Item no.	Data element	Description/notes	Data values/format
			INTR = Interstate facility PRIV = Private facility UNKN = Unknown facility Format: ANNN
59.	Degree of spread at episode	The extent of spread of cancer from its point of origin at each episode of care.	1 = Localised to tissue of origin 2 = Regional spread, adjacent organs 3 = Regional spread, regional lymph nodes 4 = Distant metastases 9 = Unknown
60.	TNM staging group	The overall TNM stage value describes the extent of an individual's cancer spread based on the primary tumour and the extent of spread in the body, according to the clinically determined values of T, N and M. T stage represents the size and extent of the primary tumour; N stage represents the extent of regional lymph node metastasis; and M stage identifies the presence or absence of distant metastasis. The TNM staging system is an internationally agreed staging classification system developed by the American Joint Committee on Cancer (AJCC) and adopted by the Union for International Cancer Control (UICC). TNM staging applies to most solid tumours but is not applicable to all tumours.	4 = I 6 = IA 7 = IA1 8 = IA2 9 = IB 10 = IB1 11 = IB2 12 = IC 13 = II 14 = IIA 15 = IIA1 16 = IIA2 17 = IIB 18 = IIC 19 = III 20 = IIIA 21 = IIIB 22 = IIIC 30 = IIIC1 31 = IIIC2 23 = IV 24 = IVA 25 = IVA1 26 = IVA2 27 = IVB 28 = IVC 29 = Occult Carcinoma 97 = Unstaged 98 = Not applicable 99 = Unknown
61.	TNM edition	TNM edition number	1 = TNM edition 1 2 = TNM edition 2 3 = TNM edition 3 4 = TNM edition 4

Item no.	Data element	Description/notes	Data values/format
			5 = TNM edition 5 6 = TNM edition 6 7 = TNM edition 7 9 = Unknown
62.	TNM staging basis	Describes the evidence for the T, N and M stage values and the overall TNM stage value. Clinical stage is based on evidence obtained prior to treatment from physical examination, imaging, endoscopy, biopsy, surgical exploration or other relevant examinations. Pathological stage is based on histological evidence acquired before treatment and is supplemented or modified by additional evidence acquired from surgery and from pathological examination. When more than one basis for staging is available, the highest stage in any one of the stages is used.	C = Clinical P = Pathological
63.	TNM staging timing	An indicator of whether the TNM staging occurred at diagnosis or not at diagnosis. Staging at diagnosis is defined within four months of the date of diagnosis.	1 = Staging at diagnosis 2 = Staging not at diagnosis
64.	TNM staging date	Date on which TNM staging was determined.	Format: DDMMYYYY
65.	MDT date	Date of consultation with a multidisciplinary team of health professionals regarding the diagnosis, treatment and care of the person with cancer. An MDT meeting in this context is a face-to-face meeting (or via video/teleconference) held at a defined time and place for the express purpose of discussing cases and deciding treatment recommendations.	Format: DDMMYYYY
66.	Performance status *(ECOG)*	A score given that appraises an individual's ability to manage activities of daily living at the time of the episode of care. Code set for performance status scale by USA Eastern Cooperative Oncology Group (ECOG). (Isselbacher et al, 1994)	0 = Fully active, able to carry on all pre-disease performance without restriction 1 = Restricted in physically strenuous activity but ambulatory, can perform light / sedentary work 2 = Ambulatory and self-caring, cannot work. Up and about > 50% of waking hours

Item no.	Data element	Description/notes	Data values/format
			3 = Only limited self-care, confined to bed or chair > 50% of waking hours 4 = Completely disabled, cannot self-care. Totally confined to bed or chair 9 = Unknown

Attachment 1 – Cancer type reporting categories (ICD-9 group)

Code	Description	Code	Description
C00	Lip	C567	Ovarian
C012	Tongue	C58	Placental
C036	Mouth	C59	Other female genital
C078	Salivary glands	C61	Prostate
C0910	Oropharyngeal	C62	Testicular
C11	Nasopharyngeal	C6360	Other male genital
C123	Hypopharyngeal	C648	Kidney
C14	Other oral cavity & pharyngeal	C67	Bladder
C15	Oesophageal	C69	Eye
C16	Stomach	C71	Brain
C17	Small intestine	C72	Central nervous system
C18	Colon	C73	Thyroid
C1921	Rectal	C745	Other endocrine glands
C22	Liver	C81	Hodgkin lymphoma
C234	Gallbladder	C82	Non-Hodgkin lymphoma
C25	Pancreatic	C8890	Multiple myeloma
C301	Nasal & sinus	C910	Acute lymphoblastic leukaemia
C32	Laryngeal	C919	Other lymphoid leukaemia
C334	Lung	C920	Acute myeloid leukaemia
C378	Other thoracic	C929	Other myeloid leukaemia
C401	Bone	C94	Other specified leukaemia
C43	Melanoma of skin	C95	Unspecified leukaemia
C44	Other skin	C98	Cancer unknown primary
C45	Mesothelioma	M95	Other lymphatic haematopoietic
C46	Kaposi's sarcoma	M96	Myelodysplasia
C479	Connective tissue, peripheral nerves	0006	Non cancer death
C50	Breast	0007	Unknown cause of death
C53	Cervical	-1	Not available
C545	Uterine		

Attachment 2 – Clinical cancer group

Code	Description	Cancer type (ICD-O-3 codes)
1	**Skin**	Lip (C00) Melanoma of the skin (C44 and M872-M879) Kaposi's sarcoma (M914)
2	**Head and neck**	Tongue (C01, C02) Mouth (C03-C06) Salivary gland (C07, C08) Oropharyngeal (C09, C10) Nasopharyngeal (C11) Hypopharyngeal (C12, C13) Other oral cavity and pharyngeal (C14) Nasal, sinus (C30, C31) Laryngeal (C32)
3	**Upper gastrointestinal**	Oesophageal (C15) Stomach (C16) Small intestinal (C17) Liver (C22) Gallbladder (C23, C24) Pancreatic (C25)
4	**Bowel**	Colon (C18) Rectal (C19-C21)
5	**Respiratory**	Lung (C33, C34) Other thoracic (C37, C38) Mesothelioma (M905)
6	**Bone and connective tissue**	Bone (C40, C41) Connective tissue, peripheral nerves (C47, C49)
7	**Breast**	Breast (C50)
8	**Urogenital**	Prostate (C61) Testicular (C62)
9	**Gynaecological**	Cervical (C53) Uterine (C54, C55) Ovarian (C56, C57.0-C57.7) Placental (C58) Other female genital (C51, C52, C57.8-C57.9)
10	**Eye**	Eye (C69)
11	**Neurological**	Brain (C71) Central nervous system (C72)
12	**Thyroid and other endocrine**	Thyroid (C73) Other endocrine glands (C74, C75)

Code	Description	Cancer type (ICD-O-3 codes)
13	**Lymphohaematopoietic**	Hodgkin lymphoma (M965-M966) Non-Hodgkin lymphoma (M959, M967-M972, M974) Multiple myeloma (M973, M976) Acute lymphoblastic leukaemia (M9821) Other lymphoid leukaemia (M9820, M9822-M9827, M994) Acute myeloid leukaemia (M9861) Other myeloid leukaemia (M9860, M9862-8, M987-M988, M9930, M9987) Other specified leukaemia (M984, M985, M989-M993) Unspecified leukaemia (M980) Myelodysplasia (M998) Other lymphoid haematopoietic (C42,C77 and M974, M975, M995-M997 [excl. M9963])
14	**Cancer unknown primary**	Other and Ill-defined digestive organs (C26) Other and Ill-defined sites in the respiratory system and intrathoracic organs (C39) Retroperitoneum and peritoneum (C48) Other and Ill-defined sites (C76) Unknown primary site (C80)
15	**Other skin**	Non melanoma skin cancer (C44 and not M872-M879)
98	**Unknown cause of death**	Non cancer death (0006) Unknown cause of death (0007)
-1	**Not available**	Not available (-1)