



Australian Government

Australian Institute of
Health and Welfare

Australian Longitudinal Study on Women's Health (ALSWH)

EO2020-3-1115

Description of Australian Cancer Database Variables

Data Integration Services Centre (DISC), July 2023

Description of Australian Cancer Database Variables

EO2019-3-1048

Contents

AIHW_PPN	3
Residence State.....	3
Sex.....	4
Registry State	4
Diagnosis_Date	4
Diagnosis_Year_Month.....	5
Diagnosis_Date_Accuracy.....	5
Diagnosis_Age.....	5
Diagnosis_Age_Group.....	5
Topography	6
Morphology.....	6
ICD10.....	6
Best_Basis_of_Diagnosis.....	6
Melanoma_Thickness	8
Breast_Tumour_Size	9
Death_Date	9
Death_Year	10
Death_Age.....	10
Death_Age_Group	10
Underlying_Cause_of_Death	11
Birth_Date.....	11
Birth_Date_Accuracy	12
SA2_2011	12
SA2_2016	12

Description of Australian Cancer Database Variables

EO2019-3-1048

The ACD results are found in the following files:

eo202031115_acd_content_july23.sas7bdat

The number of records in each is 8,073.

The following is a description of the fields:

AIHW_PPN

Type: Character, length 13.

Meaning: A unique identifier for each person in the cohort.

Meaning of null value: Impossible.

Residence State

Type: Character, length 3.

Meaning: State or territory of the usual residence of the person at the time of their diagnosis.

Meaning of null value: Estimated data.

Value	Meaning
ACT	Australian Capital Territory
NSW	New South Wales
NT	Northern Territory
OT	Christmas Island, Cocos (Keeling) Islands, Jervis Bay Territory or Norfolk Island
QLD	Queensland
SA	South Australia
TAS	Tasmania
VIC	Victoria
WA	Western Australia

Description of Australian Cancer Database Variables

EO2019-3-1048

Sex

Type: Numeric, integer.

Meaning: Sex of the person.

Meaning of null value: Impossible.

Value	Meaning
1	Male
2	Female
0	Intersex/non-binary/other

Registry State

Type: Character, length 3.

Meaning: Cancer registry in which the case is registered.

Meaning of null value: Impossible.

Note 1: As shown in the table below, some registries cover more than one jurisdiction. In such cases the local_person_id numbers are shared between the jurisdictions. For example, local person id 123 in state ACT is the same person as local person id 123 in state NSW.

Note 2: Norfolk Island cases can be registered in NSW or QLD. If a person has tumours registered in both, their local person ids will not be the same.

Value	Meaning
NSW	NSW+ACT. Includes all cases from Jervis Bay Territory and some from Norfolk Island
NT	NT
QLD	Qld. Includes some cases from Norfolk Island
SA	SA
TAS	Tas
VIC	Vic
WA	WA. Includes all cases from Christmas Island and Cocos (Keeling) Islands

Diagnosis_Date

Type: Date.

Meaning: Date on which the tumour was diagnosed.

Meaning of null value: Not authorised for that particular state (for ACT in this case).

Note: Diagnosis_date is a full date even if diagnosis_day and/or diagnosis_month are unknown (null). If any parts are imputed, this is recorded in the field

Description of Australian Cancer Database Variables

EO2019-3-1048

diagnosis_date_accuracy. In the case of day imputation, the SAS code “day(diagnosis_date)” will give the imputed day whereas diagnosis_day will be null, and similarly for month.

Diagnosis_Year_Month

Type: Type: Character, length7.

Meaning: Year and Month (1-12) in which the tumour was diagnosed.

Diagnosis_Date_Accuracy

Type: Character, length 3.

Meaning: An indicator of the accuracy of diagnosis_date.

Meaning of null value: Accurate date or estimated data.

See the entry for birth_date_accuracy.

Diagnosis_Age

Type: Numeric, integer.

Meaning: The age, in completed whole years, of the person on the day of their diagnosis.

Meaning of null value: Estimated data.

Note: Diagnosis_age is defined to be the value of diagnosis_age_exact rounded down to the nearest integer.

Usage: Do not use this variable for computing mean or median age at diagnosis. Use diagnosis_age_exact.

Special value	Meaning
999	Unknown age at diagnosis

Diagnosis_Age_Group

Type: Numeric, integer.

Meaning: The age group of the person on the day of their diagnosis.

Meaning of null value: Impossible.

Note: The age groups, numbered 1 to 19, are the 5-year age groups 0-4, ..., 85-89 and 90+.

Special value	Meaning
99	Unknown age at diagnosis

Description of Australian Cancer Database Variables

EO2019-3-1048

Topography

Type: Character, length 4.

Meaning: ICD-O-3.2 topography code of the tumour.

Meaning of null value: Impossible.

Morphology

Type: Character, length 6.

Meaning: ICD-O-3.2 morphology code of the tumour.

Meaning of null value: Impossible.

Note: The morphology code is simply the concatenation of the histology code, a slash, and the behaviour code, e.g. if histology = 8500 and behaviour = 3 then morphology = 8500/3.

ICD10

Type: Character, length 4.

Meaning: The ICD-10 (2019 version) disease code of the tumour.

Meaning of null value: Impossible.

Best_Basis_of_Diagnosis

Type: Numeric, integer.

Meaning: The most valid basis of diagnosis of the tumour.

Meaning of null value: Impossible.

Note: A tumour can be diagnosed by a number of different methods, each of which has a degree of reliability associated with it. The value contained in this field represents the

Description of Australian Cancer Database Variables

EO2019-3-1048

method of the most reliable method used, using the hierarchy 7 > 6 > ... > 1 > 0 (see table below).

Value	Meaning
0	Death certificate only (DCO): information provided is from a death certificate. No other information has been found regarding the neoplasm
1	Clinical only: diagnosis made before death, but without any of the following (codes 2–8)
2	Clinical investigation: any diagnostic technique, including x-ray, endoscopy, imaging, ultrasound, exploratory surgery (e.g. laparotomy) and autopsy, without a tissue diagnosis
4	Specific tumour markers: including biochemical and/or immunological markers that are specific for a tumour site
5	Cytology: examination of cells from a primary or secondary site, including fluids aspirated by endoscopy or needle; also includes the microscopic examination of peripheral blood and bone marrow aspirates
6	Histology of metastasis: histological examination of tissue from a metastasis, including autopsy specimens
7	Histology of a primary tumour: histological examination of tissue from primary tumour, however obtained, including all cutting techniques and bone marrow biopsies; also includes autopsy specimens of primary tumour
8	Histology: either unknown whether of primary or metastatic site, or not specified
9	Unknown

Description of Australian Cancer Database Variables

EO2019-3-1048

Melanoma_Thickness

Type: Numeric, float.

Meaning: The Breslow thickness of a malignant melanoma of the skin, measured in millimetres.

Meaning of null value: The tumour is not a malignant melanoma of the skin, or estimated data.

Note 1: The proportion of records with a known value (less than 990) varies by jurisdiction and year. If you need to calculate statistics on this variable such as the mean or median, it is important to do some exploratory analysis first in order to determine if there are “enough” records with a known value.

Note 2: If you are analysing this variable after having extracted records for which the ICD-10 code is C43, you may encounter some records whose Breslow thickness is null. Here is the explanation for this phenomenon. Melanomas with a topography code of C80.9 represent melanomas of unknown primary site. As such, their Breslow thickness is null. However, according to the ICD-10 coding rules, these cancers are reported under ICD-10 code C43.9 (“melanoma, NOS”), i.e. they are reported *within ICD-10* as if they are melanomas of the skin, not cancers of unknown primary site.

Note: **xx. This variable needs some cleaning up and some extra checks in Validata.** For invasive melanoma of the skin all values should be positive but some are null (all registries have some such cases). For invasive melanoma of other sites all values should be null or positive but WA has some negative values. For melanoma in situ of any site all values should be blank but some are non-blank for TAS, VIC and WA. For non-melanoma of any site all values should be blank but there are some non-blank values for VIC and WA.

Special value	Meaning
Null	Tumour is not a malignant melanoma
9.89	For SA and NT records the value 9.89 means greater than or equal to 9.89
9.99	For Vic records the value 9.99 means greater than or equal to 9.99
990	Stated to be microinvasive but no size given. This code is not used by any registries!
999	Unknown

Description of Australian Cancer Database Variables

EO2019-3-1048

Breast_Tumour_Size

Type: Numeric, float.

Meaning: The size of an invasive breast cancer, in millimetres.

Meaning of null value: The tumour is not an invasive breast cancer, or estimated data.

Note: The proportion of records with a known value (less than 990) varies by jurisdiction and year. If you need to calculate statistics on this variable such as the mean or median, it is important to do some exploratory analysis first in order to determine if there are “enough” records with a known value.

Note: **xx. This variable needs some cleaning up and some extra checks in Validata.** For invasive breast cancer (C50) all values should be positive but some are null (all registries have some such cases). For breast carcinoma in situ (D05) all values should be blank. However, there are some non-blank values for TAS and WA. For all other neoplasms all values should be blank but there are some non-blank values for TAS, VIC and WA.

Special value	Meaning
Null	Tumour is not an invasive breast cancer
3	In all NT records and in SA records diagnosed 1982–2010 a value of 3 means less than or equal to 3.
990	Stated to be microinvasive but no size given. Not all registries have a separate domain value for this concept.
995	Occult primary, i.e. cancer cells were not found in the breast tissue excised but were found in at least one lymph node excised. Not all registries have a separate domain value for this concept. In fact, this value does not occur at all in the 2019 ACD. It did occur in the 2017 ACD for some VIC cases.
999	Unknown or not applicable (e.g. phyllodes tumour)

Death_Date

Type: Date.

Meaning: Date on which the person died.

Meaning of null value: Person is still alive, or not known to be dead, or estimated data, or only the death year is approved to be released (for ACT in this case).

Description of Australian Cancer Database Variables

EO2019-3-1048

Death_Year

Type: Numeric, integer.

Meaning: Year in which the person died.

Meaning of null value: Person is still alive, or not known to be dead, or estimated data.

Death_Age

Type: Numeric, integer.

Meaning: The age, in completed whole years, of the person at the time of their death.

Meaning of null value: Person is still alive, or not known to be dead, or estimated data.

Note: Death_age is defined to be the value of death_age_exact rounded down to the nearest integer.

Usage: Do not use this variable for computing mean or median age at death. Use death_age_exact.

Special value	Meaning
Null	Person is still alive, or not known to be dead
999	Unknown age at death

Death_Age_Group

Type: Numeric, integer.

Meaning: The age group of the person on the day of their death.

Meaning of null value: Person is still alive, or not known to be dead, or estimated data.

Note: The age groups, numbered 1 to 19, are the 5-year age groups 0-4, ..., 85-89 and 90+.

Special value	Meaning
99	Unknown age at death

Description of Australian Cancer Database Variables

EO2019-3-1048

Underlying_Cause_of_Death

Type: Character, length 4.

Meaning: The consolidated cause of death variable to use for reporting from the ACD.

Meaning of null value: Person is still alive, or not known to be dead, or estimated data.

Note 1: See the entries for abs_cod, cod_flag and std_state_cod_icd10.

Note 2: A person might have several different sources of cause of death (COD). Each cancer registry in which the person is registered will have a cancer registry-derived COD for them, which is called the CR COD. Its value is stored in std_state_cod_icd10. Note that different cancer registries might allocate different CR CODs to the person. Furthermore, the ABS will have a COD for them, called the ABS COD, which is obtained via linkage of the ACD to the NDI. Its value is stored in abs_cod. Note that the ABS COD might not be the same as the CR COD(s). A business rule is applied so that each national person id has a unique COD, even if there are multiple CODs available. The rule is as follows.

1. Use the CR COD of the most recent diagnosis unless it is blank or ZUNK, in which case use the CR COD of the second most recent diagnosis unless it is blank or ZUNK, and so on. In other words, keep working backwards through the diagnoses until the first CR COD that is not blank or ZUNK is found.
2. If all of the CR CODs are blank or ZUNK, use the ABS COD. However, if the ABS COD is blank, assign ZUNK as the cause_of_death rather than blank. Also, if the ABS COD is not cancer, assign ZNCD rather than the specific code. Finally, if the ABS COD is C77-C80 or C97, assign C80.9.

Birth_Date

Type: Date.

Meaning: Date on which the person was born.

Meaning of null value: Person's date of birth is unknown or estimated data.

Note: Provided that birth_year is not null, birth_date will be a full date even if birth_day and/or birth_month are unknown (null). If any parts are imputed, this is recorded in the field birth_date_accuracy. In the case of day imputation, the SAS fragment "day(birth_date)" will give the imputed day whereas birth_day will be null, and similarly for month.

Description of Australian Cancer Database Variables

EO2019-3-1048

Birth_Date_Accuracy

Type: Character, length 3.

Meaning: An indicator of the accuracy of birth_date.

Meaning of null value: Accurate date or estimated data.

The date accuracy indicator consists of up to three letters. "D" and "d" refer to the day, "M" and "m" refer to the month, and "Y" and "y" refer to the year. If a letter is present it means the component was estimated. If a letter is absent it means the component is not recorded as having been estimated; nevertheless, it is possible that it was estimated by the registry. If a letter is upper case it means that the estimate was made by the registry. If a letter is lower case it means the registry supplied a blank value and the AIHW estimated the component.

Special value: A null value means that the registry has recorded the date as accurate. However, it is possible that the date is partly or fully estimated and that the registry simply didn't record this fact. A null value also occurs with all estimated data.

Special value: UUU means that the date is entirely unknown.

Note: QLD only records a yes/no as to whether a date is estimated, i.e. they don't record which components were estimated and which are accurate. The date accuracy indicator represents this condition with DMY. It means that at least one component was estimated without specifying which one(s). In other registries DMY would mean that all three components were estimated.

SA2_2011

Type: Numeric, integer.

Meaning: 9-digit Statistical Area Level 2 (SA2), in the 2011 ASGS, of the usual residence of the person at the time of their diagnosis.

Meaning of null value: Impossible.

Special value: 0 (zero) means the SA2 is unknown or the person had no usual address.

SA2_2016

Type: Numeric, integer.

Meaning: 9-digit Statistical Area Level 2 (SA2), in the 2016 ASGS, of the usual residence of the person at the time of their diagnosis.

Meaning of null value: Impossible.

Special value: 0 (zero) means the SA2 is unknown or the person had no usual address.