

Process for confidentialising the 2022–2023 NDSHS dataset

The approach taken to confidentialise the 2022–2023 National Drug Strategy Household Survey (NDSHS) unit record file has been conducted in line with the process employed for the 2019 NDSHS, to ensure that the dataset is confidential and that the requirements of the AIHW Ethics Committee have been met.

AIHW Ethics Committee requirements

The AIHW Ethics Committee requirements for publication of the NDSHS unit record data stipulate that transformations are made to the data such that they cannot be used to identify or re-identify individuals. These transformations include:

- The original fully open-ended responses to questions ZZ7 and ZZ8 are not included on the CURF. ANZSCO (Australian and New Zealand Standard Classification of Occupations) and ANZSIC (Australian and New Zealand Standard Industrial Classification) codes were used as the preferred code frames for these questions (ZZ7 and ZZ8 respectively), aggregated to 2 digits.
- Statistical Area Level 1 (SA1) and Statistical Area Level 2 (SA2) have been de-identified (but are hierarchical).
- Postcode has been removed.
- Australian Statistical Geography Standard (ASGS) remoteness areas have been aggregated and there are two variables available on the CURF.
 - ASGS4 – Use this variable for Qld, SA, WA, and NT. The value labels are:
 - Major cities
 - Inner Regional
 - Outer regional
 - Remote/very remote
 - ASGS3 – Use this variable for NSW, Vic, Tas and ACT. This variable can also be used for all states and territories if comparing between state and territories (national remoteness analysis). The value labels are:
 - Major cities (*Inner regional is grouped with major cities for ACT)
 - Inner regional
 - Outer regional/remote/very remote

If you require access to the original variables, you may need to submit an Ethics application. See below for details.

Confidentiality requirements

All potentially identifiable variables were assessed to ensure that the dataset is sufficiently de-identified. These variables included all demographic variables, sensitive questions and health variables that may potentially identify an individual. No adjustments have been made to any drug use or attitudes variables. The process and additional steps required for confidentialising the unit record file to ensure that the data set is sufficiently de-identified are outlined below.

1. The first step involved determining a minimum threshold for the count for any variable. This threshold was set at 2% of the total sample size (or a minimum count of 434). All variables where the count of individual responses total greater than 2% are included on the CURF, unless further sensitivities were considered to increase the risk of re-identification.
2. The second step required aggregating response categories for variables where the count of individual responses was less than or equal to 2%. Response codes were grouped into meaningful categories and then reassessed. If aggregated response codes were greater than 2% then the aggregated responses are included on the CURF.
3. If aggregated responses still totalled less than 2%, further analysis was conducted and an assessment was made as to whether the variable is of critical importance for analysis, that is, if the analysis of drug use patterns is sufficiently affected by this variable.
4. If this was not the case, then the variable was removed from the general release CURF.
5. If the variable was deemed to be important for analysis, it has either been:
 - a. included on the CURF but requires specific condition of use to be met;
 - b. moved to a restricted access file and researchers will need to apply for special access to these variables and explain why the variable is required for their analysis; or
 - c. moved to a special purpose dataset and researchers will need to apply for special access to these variables.

Implications for individual questions

Response categories greater than minimum threshold

These variables were assessed and are included on the dataset in a format consistent with the questionnaire.

- DEM6B – Dependent children in the household

Aggregated responses greater than the minimum threshold

Individual response codes were grouped into meaningful categories and the grouped categories are included on the dataset (rather than the original response codes in the questionnaire).

Questionnaire number	New variable name	Response categories
DEM1B	Gender	1 'Man or male'

		2 'Woman or female' 3 'Non-binary, I use another term, or not answered' -1 'Missing'
DEM2	CURF_Age	Grouped those aged 80 or over in the mean age for this group (84.05)
DEM4	Indigeneity	1 'First Nations people' 2 'Non-Indigenous people' -1 'Missing'
DEM3	MaritalStatus	1 'Never married' 2 'Divorced/Separated/Widowed' 3 'Married/de facto' -1 'Missing'
DEM5	Household	1 'Single with dependents' 2 'Couple with dependents' 3 'Parents with non-dependent children' 4 'Singles without children' 5 'Couple without children' 6 'Other' -1 'Missing'
DEM6A	NumberPpl	1 '1' 2 '2' 3 '3' 4 '4' 5 '5 or more'
DEM6C	NumberDepen	1 '1' 2 '2' 3 '3 or more' 4 'No dependent children in household'
DEM7_01	Children0to2	1 'Yes dependent children aged between 0-2' 2 'No children aged between 0-2' 3 'No dependent children in household'
DEM7_02 DEM7_03 DEM7_04	Children3to14	1 'Dependent children aged 3 to 14' 2 'No children aged 3 to 14' 3 'No dependent children in household'

DEM7_05		
DEM7_06	Children15p	1 'Dependent children aged 15 or older' 2 'No children aged 15 or older' 3 'No dependent children in household'
B4_01 B4_02	Diabetes	1 'Yes, have been diagnosed and/or treated' 2 'No, have not been diagnosed or treated'
B4_03	Heart_disease	1 'Yes, have been diagnosed and/or treated' 2 'No, have not been diagnosed or treated'
B4_04	Hypertension	1 'Yes, have been diagnosed and/or treated' 2 'No, have not been diagnosed or treated'
B4_05	LowIron	1 'Yes, have been diagnosed and/or treated' 2 'No, have not been diagnosed or treated'
B4_06	Asthma	1 'Yes, have been diagnosed and/or treated' 2 'No, have not been diagnosed or treated'
B4_07	Depression	1 'Yes, have been diagnosed and/or treated' 2 'No, have not been diagnosed or treated'
B4_08	Anxiety	1 'Yes, have been diagnosed and/or treated' 2 'No, have not been diagnosed or treated'
B4_09 B4_10 B4_11 B4_12	Mental_other	1 'Yes, have been diagnosed and/or treated' 2 'No, have not been diagnosed or treated'
B4_07	Mental_health	Treated and diagnosed for all mental health conditions have been grouped together.

B4_08		1 'Yes, have been diagnosed and/or treated'
B4_09		2 'No, have not been diagnosed or treated'
B4_10		
B4_11		
B4_12		
B4_14	Chronic	1 'Yes, have been diagnosed and/or treated' 2 'No, have not been diagnosed or treated'
B4_16	Cancer	1 'Yes, have been diagnosed and/or treated' 2 'No, have not been diagnosed or treated'
B4_13	OtherIllness	1 'Yes, have been diagnosed and/or treated'
B4_15		2 'No, have not been diagnosed or treated'
B4_17		
Y2A_02, Y2A_03, Y2A_04, Y2A_05	Y2A_Relative	The person under the influence of alcohol who verbally abused you in Y1 was either a parent, child, sibling or other relative of the respondent. 1 'Yes' 2 'No'
Y2A_01, Y2A_07, Y2A_08	Y2A_Partner	The person under the influence of alcohol who verbally abused you in Y1 was either a current or ex-spouse, partner or boy/girlfriend of the respondent 1 'Yes' 2 'No'
Y2A_09, Y2A_10, Y2A_11, Y2A_06	Y2A_Known	The person under the influence of alcohol who verbally abused you in Y1 was either a friend, work/school mate, flat/house mate or other person known to the respondent 1 'Yes' 2 'No'
Y2C_02, Y2C_03, Y2C_04, Y2C_05	Y2C_Relative	The person under the influence of alcohol who put you in fear in Y1 was either a parent, child, sibling or other relative of the respondent. 1 'Yes' 2 'No'
Y2C_01, Y2C_07, Y2C_08	Y2C_Partner	The person under the influence of alcohol who put you in fear in Y1 was either a current or ex-spouse, partner or boy/girlfriend of the

		<p>respondent</p> <p>1 'Yes'</p> <p>2 'No'</p>
Y2C_09, Y2C_10, Y2C_11, Y2C_06	Y2C_Known	<p>The person under the influence of alcohol who put you in fear in Y1 was either a friend, work/school mate, flat/house mate or other person known to the respondent</p> <p>1 'Yes'</p> <p>2 'No'</p>
Y10A_02, Y10A_03, Y10A_04, Y10A_05	Y10A_Relative	<p>The person under the influence of illicit drugs who verbally abused you in Y9 was either a parent, child, sibling or other relative of the respondent.</p> <p>1 'Yes'</p> <p>2 'No'</p>
Y10A_01, Y10A_07, Y10A_08	Y10A_Partner	<p>The person under the influence of illicit drugs who verbally abused you in Y9 was either a current or ex-spouse, partner or boy/girlfriend of the respondent</p> <p>1 'Yes'</p> <p>2 'No'</p>
Y10A_09, Y10A_10, Y10A_11, Y10A_06	Y10A_Known	<p>The person under the influence of alcohol who verbally abused you in Y9 was either a friend, work/school mate, flat/house mate or other person known to the respondent</p> <p>1 'Yes'</p> <p>2 'No'</p>
Y10C_02, Y10C_03, Y10C_04, Y10C_05	Y10C_Relative	<p>The person under the influence of illicit drugs who put you in fear in Y9 was either a parent, child, sibling or other relative of the respondent.</p> <p>1 'Yes'</p> <p>2 'No'</p>
Y10C_01, Y10C_07, Y10C_08	Y10C_Partner	<p>The person under the influence of illicit drugs who put you in fear in Y9 was either a current or ex-spouse, partner or boy/girlfriend of the respondent</p> <p>1 'Yes'</p> <p>2 'No'</p>
Y10C_09, Y10C_10, Y10C_11, Y10C_06	Y10C_Known	<p>The person under the influence of alcohol who put you in fear in Y9 was either a friend, work/school mate, flat/house mate or other person known to the respondent</p> <p>1 'Yes'</p> <p>2 'No'</p>
Z5A_01, Z5A_02, Z5A_03	Pregnancy	<p>1 'Pregnant/Breastfeeding'</p> <p>2 'Breastfeeding only'</p>

		3 'Not pregnant or breastfeeding' -1 'Missing'.
Z6A	WeeksPreg	1 '<12 weeks' 2 '13-28 weeks' 3 '29-42 weeks' 4 'Not currently pregnant'
ZZ1	COB	Country of Birth grouped by 2016 Standard Australian Classification of Countries
ZZ3	MainLanguage	1 'English' 2 'Language other than English' -1 'Missing'
ZZ4	Sexuality	Grouped categories 1 'Straight (heterosexual)' 2 'Gay or lesbian, or bisexual' 3 'I use a different term/Don't know/Not answered'
ZZ5A	EmploymentStatus	1 'Currently employed' 2 'Student' 3 'Unemployed/Looking for work' 4 'Solely engaged in home duties' 5 'Retired or on a pension' 6 'Unable to work' 7 'Other' -1 'Missing'
ZZ5B	Employment2	1 'Not in labour force' 2 'Unemployed/Looking for work' 3 'Currently employed' -1 'Missing'
ZZ7	ANZSCO_2	ANZCO codes as described above
ZZ8	ANZSIC_D	ANZSIC codes as described above
ZZ9	ZZ9	Combined did not go to school, year 6 or below and year 7.
ZZ12	HighestQualification	1 'Certificate I or II' 2 'Certificate III or IV' 3 'Associate diploma' 4 'Undergraduate diploma' 5 'Bachelor degree' 6 'Masters, postgrad, or doctorate' -1 'Missing'

ZZ13	PersIncome	All original income ranges included but nil income and negative income are grouped together.
ZZ14	HHIncome	Amounts from \$1 to \$20,799 are grouped together. Nil income and negative income are grouped together.

Removed variables

Variables that did not meet above criteria or removed due to confidentiality concerns

Aggregated responses that are still less than the minimum threshold have been removed from the general release CURF (except for those variables identified as important for analysis but require specific conditions of use). The following variables have been removed from the general release dataset but may be available as a special request file:

- E13C – Required medical attention/admission to hospital while under the influence of alcohol.
- W4 – Required medical attention/admission to hospital while under the influence of illicit drugs.
- Y1B – Physically abused by someone under the influence of alcohol.
- Y2A – Persons responsible for alcohol-related incidents involving verbal abuse.
- Y2B – Persons responsible for alcohol-related incidents involving physical abuse.
- Y2C – Persons responsible for alcohol-related incidents involving put you in fear.
- Y3A – Places where alcohol-related incidents involving verbal abuse happened.
- Y3B – Places where alcohol-related incidents involving physical abuse happened.
- Y3C – Places where alcohol-related incidents involving put you in fear happened.
- Y4 – Most serious physical injury sustained as result of alcohol-related incident.
- Y8 – Alcohol-related incidents involved sexual abuse.
- Y9B – Physically abused by someone under the influence of illicit drugs.
- Y10A – Drug-related incidents involving verbal abuse.
- Y10B – Drug-related incidents involving physical abuse.
- Y10C – Drug-related incidents involving put you in fear.
- Y11A – Places where drug-related incidents involving verbal abuse happened.
- Y11B – Places where drug-related incidents involving physical abuse happened.
- Y11C – Places where drug-related incidents involving put you in fear happened.
- Y12 – Most serious physical injury sustained as result of drug-related incident.
- Y16 – Drug-related incidents involved sexual abuse.
- Z6A – Date fell pregnant.

- Z7 – Currently pregnant, breastfeeding or neither.

Variables removed due to questionnaire design changes

As in 2019, two alcohol-related questions (E15 and E16) were presented in an alternative format to 50% of online respondents. The goal of this approach was to assess whether this improved the quality of respondents' answers (see the [Technical Information](#) for more information).

This change to the presentation of the questions resulted in substantial differences in the pattern of people's responses. To preserve comparability with previous survey waves, the 2022–2023 CURF has removed responses where the alternative format was presented for the following variables:

- E15
- E16
- Total_guideline
- Ten_per_week
- Four_in_day_daily
- Four_in_day_weekly
- Four_in_day_monthly
- Four_in_day_yearly
- Four_in_day
- Eleven_day
- Eleven_week
- Eleven_month
- Eleven_year

Any analysis using these variables should use an alternative weighting scheme (see the Weighting section below for more information). Full responses to questions E15 and E16 (including those who answered an alternative format) are available on request for researchers who are interested in undertaking a specific research project that aims to assess which question format captures better alcohol data.

Variables with specific conditions of use

These variables did not meet minimum threshold, either alone or cross tabulated with other demographic variables, but are deemed to be important for analysis so have been included on the general release CURF. They **must not** be crossed tabbed with other demographic or geographic variables (age groups are an exception to this rule). If researchers wish to do this, they must seek approval from the AIHW.

Questionnaire number	Variable name	Value labels (original or grouped categories)
----------------------	---------------	---

DEM2	CURF_Age	Original (except people aged 80 or older have been perturbed to mean age of 84.05)
DEM1B	Gender	Original responses were categorised into three groups: 1 'Man or male' 2 'Woman or female' 3 'Non-binary, I use another term, or not answered' -1 'Missing'
ZZ4	Sexuality	Grouped categories: 1 'Heterosexual' 2 'Homosexual/bisexual' 3 'Other' -1 'Missing'
ZZ2	ArrivalYear	Grouped categories: 1 '5 years or less' 2 '6–10 years' 3 '11–19 years' 4 '20 years or more' 5 'Not applicable – will be in Australia for less than a year' -1 'Missing, Invalid response'
Z5A_01 Z5A_02 Z5A_03	Pregnancy	Original responses were grouped: 1 'Pregnant/Breastfeeding' 2 'Breastfeeding only' 3 'Not pregnant or breastfeeding' -1 'Missing'
ZZ1 ZZ3 ZZ3A ZZ3B	CALD_flag	Derived variable using: 1 'Culturally and Linguistically Diverse background' 2 'Not CALD background'
ZZ3	MainLanguage	Original responses were grouped: 1 'English' 2 'Language other than English' -1 'Missing'
ZZ3A	FirstLanguage	Original responses were grouped: 1 'English' 2 'Language other than English' -2 'Missing'
ZZ8	Industry	ANZSIC (grouped as described above)
ZZ7	Occupation	ANZCO (grouped as described above)

Special request variables

Some variables have been removed from the general release CURF but are available as additional data sets on the ADA website—either as a special purpose dataset or restricted access file (see table below). Researchers may request access to these variables in addition to the general release CURF, depending on the project.

2022–2023 variable/question number	Description of variable	Special access type
DEM1A and DEM1B	Sex at birth and gender (all responses).	Restricted access
Y1B Y2B Y2C Y3A Y3B Y3C Y4 Y8	Physical abuse data for incidents related to alcohol, Follow-up questions for all alcohol-related incidents.	Restricted access
Y9B Y10A Y10B Y10C Y11A Y11B Y11C Y12 Y16	Physical abuse data for incidents related to illicit drugs, Follow-up questions for all illicit drug-related incidents	Restricted access
Z4	Participation in treatment programs (some researchers may have additional access to data/information that could make this potentially identifiable.	Restricted access
ZZ4	Sexual orientation (all responses).	Restricted access
ZZ8A	Service in the Australian Defence Force	Restricted access
E13C and W4	Required medical attention/admission to hospital while under the influence of alcohol/illicit drugs	Special purpose data
E15 E16	Full responses to questions E15 and E16 (including the online respondents who answered an alternative format)	Special purpose data

DEM1A DEM1B	GenderExperience	Derived variable categorised into two groups: 1 'Cisgender' 2 'Trans and gender diverse' -1 'Missing or Inadequately described'
----------------	------------------	--

Variables that require approval from the AIHW ethics committee

Several variables may only be made available through an ethics application process. These include:

- Full breakdown of sex and gender variables.
- Geographic variables, including Statistical Areas and postcodes.

Information on making an ethics application can be found [online](#).

Respondents removed from the CURF

In previously released NDSHS CURFs, respondents who could not be adequately de-identified within the data were removed. In 2019, this included respondents from remote First Nations communities, and people whose sex was recorded as something other than male or female. These de-identification categories were not required in the 2022–2023 CURF, as the survey was not distributed in remote communities where English was not the primary spoken language, and updates to the sex and gender variables, and the weighting scheme, allowed for missing values in the dataset. More information can be found in the [technical notes section of the online report](#).

As a result, weighted estimates generated using the 2022–2023 NDSHS CURF should match those published in the [online data tables](#). This is different to previous years, where results may only match online published data for breakdowns by state and territory and gender, such as Table 9b.5 in 2022–2023.

Weighting

Respondents within each stratum were assigned weights to overcome imbalances arising in the design and execution of the sampling, as well as differences in response rates and to correct for over-sampled strata. The main weighting took into account geographical stratification, household size, age and gender.

The CURF includes several weights that should be used depending on the questions being used for analysis, and whether the unit of interest is the survey respondent or the entire household.

Weight	Purpose
Weight_7	This is the respondent-level weight and is the primary weight that should be used for analysis. Any results that examine respondents

	<p>(for example: smoking status, drinking status, use of illicit drugs or attitudes) should use weight_7.</p> <p>The only results that should not use weight_7 are analyses examining the proportion of households, or analysis involving the amount of alcohol consumed.</p>
Weight_7_alc_amount	<p>This is also a respondent-level weight and should be used in any analysis that references the amount of alcohol that each respondent consumed, based on question E16. This is due to people who answered the non-grid format of E16 being excluded from analysis, as detailed in the Technical Information.</p> <p>This includes any analysis of variables based on the Australian guidelines to reduce health risks from drinking alcohol.</p>
Weight_8	<p>This is the household-level weight and should be used in any analysis where the household is the unit of interest. For example, questions involving the proportion of people who are exposed to tobacco smoke in the household should use weight_8.</p>