

Process for Confidentialising the 2010 NDSHS Dataset

The AIHW are working to achieve greater consistency in confidentiality and data quality assessment practice across the organisation. The approach taken to confidentialise the 2010 National Drug Strategy Household Survey (NDSHS) unit record file has been investigated 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 imposed two conditions on 2010 data relating to confidentiality; that the occupation coding be truncated to two digits, and that no geographic area of fewer than 50,000 persons be identifiable. The result of these two conditions means that:

- The original fully open-ended responses to questions ZZ11 and ZZ12 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 codeframes for these questions (ZZ12 and ZZ11 respectively), aggregated to 2 digits.
- Statistical Local Area (SLA) and Census Collection District (CCD) have been de-identified (but hierarchical).
- Postcode has been removed.
- Australian Standard Geographical Classification (ASGC) remoteness areas have been aggregated and there are two variables available on the CURF.
 - ASGCRemoteness1 – Use this variable for Qld, SA, WA, NT, ACT*. Value labels:
 - Value labels:
 - Major cities (*Inner regional is grouped with major cities for ACT)
 - Inner Regional
 - Outer regional
 - Remote/very remote
 - ASGCRemoteness2 – Use this variable for NSW, Vic, Tas. This variable can also be used for all states and territories if comparing between state and territories (national remoteness analysis).
 - Value labels:
 - Major cities (*Inner regional is grouped with major cities for ACT)
 - Inner regional
 - Outer regional/remote/very remote

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 with potential risk. 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. This threshold was set at 2% (or sample size of 532). All variables with individual responses of greater than 2% are included on the CURF.

2. The second step required aggregating response categories for variables with individual responses of less than 2%. Response codes were grouped into meaningful categories and then reassessed to determine if categories still fell below the 2% threshold. If aggregated response codes were greater than 2% then the aggregated responses are included on the CURF.
3. If still less than 2%, further analysis was conducted and an assessment made of whether the variable is of critical importance for analysis, that is, do drug use patterns differ by this subpopulation.
4. If this was not the case, then the variable was removed from the CURF.
5. If the variable is deemed to be important for analysis, the variable is included on the CURF but may require specific conditions of use to be met (see conditions of use later in this document and on the separate conditions of use form).

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.

- ZZ3 - Marital status
- ZZ14 - Completed trade or other qualification

Aggregated responses greater than 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	Variable name	Grouped Categories
B2	Height	<ol style="list-style-type: none"> 1. <155cm 2. 155-164cm 3. 165-174cm 4. 175-184cm 5. 185+cm
B3	Weight	<ol style="list-style-type: none"> 1. <50kg 2. 50-54kg 3. 55-59kg 4. 60-64kg 5. 65-69kg 6. 70-74kg 7. 75-79kg 8. 80-84kg 9. 85-89kg 10. 90-94kg 11. 95-99kg 12. 100-109kg 13. 110kg+
B6	Health Conditions	Treated and diagnosed have been grouped together
B6_01	Insulin dependent diabetes	Diabetes
B6_02	Non-insulin dependent diabetes	
B6_03	Heart disease	Heart disease

B6_04	Hypertension	Hypertension
B6_05	Low iron	Low iron
B6_06	Asthma	Asthma
B6_07	Depression	Mental disorder
B6_08	Anxiety Disorder	
B6_09	Schizophrenia	
B6_10	Bi-polar disorder	
B6_11	Other form of psychosis	
B6_12	An eating disorder	
B6_13	A sexually transmitted infection	Other
B6_14	Hepatitis B or C	
B6_16A	Other Illness	
Y4	Most serious physical injury you sustained as a result of the alcohol-related incident(s)	<ol style="list-style-type: none"> 1. Bruising and abrasion 2. All other Injuries
Z6b	Number of weeks pregnant when pregnancy confirm	<ol style="list-style-type: none"> 1. 1-5 weeks 2. 6 or more weeks
ZZ7	Main Language	<ol style="list-style-type: none"> 1. English 2. Other
ZZ8	Other Language	<ol style="list-style-type: none"> 1. English 2. Other
ZZ9a	Main Employment	<ol style="list-style-type: none"> 1. Self-employed 2. Employed 3. Unemployed/looking for work/unable to work 4. Home duties 5. Student 6. Retired 7. Other
ZZ9b	Other Employment	<ol style="list-style-type: none"> 1. Other employment 2. No other employment
ZZ15	Highest qualification obtained	<ol style="list-style-type: none"> 1. Trade certificate 2. Non-trade certificate 3. Associate or undergraduate diploma 4. Bachelor degree 5. Masters, postgrad, or doctorate
ZZ16	Personal Income	All original income ranges included but nil income and negative income are grouped together
ZZ17	Household Income	All original income ranges included except for bottom two ranges which are grouped together (\$1-\$12,999) and nil income and negative income are grouped together
ZZ18a	Number people aged 12 over in household	<ol style="list-style-type: none"> 1. One 2. Two 3. Three 4. Four 5. Five or more
ZZ18c	Number of children parent or guardian for	<ol style="list-style-type: none"> 1. One 2. Two 3. three or more
ZZ19	Age of dependent children	<ol style="list-style-type: none"> 1. Dependent children aged 0-2 years 2. Dependent children aged over 2.

ZZ20	Household structure	<ol style="list-style-type: none"> 1. Single with dependents 2. Couple with dependents 3. Parents with non-dependents 4. Singles without children 5. Couples without children 6. Other
------	---------------------	--

Removed variables

Did not meet above criteria, aggregated response codes are still less than minimum threshold and therefore variable must be removed from dataset.

- Y8 – alcohol-related incidents involved sexual abuse
- Y12 – most serious physical injury sustained as result of drug-related incident
- Y16 – drug-related incidents involved sexual abuse
- Z6a – date fell pregnant
- ZZ6b – year arrived in Australia

Variables with specific conditions of use

These variables did not meet minimum threshold but are deemed to be important for analysis:

Questionnaire Number	Variable name	Value labels (original or grouped categories)
ZZ2	Age	Original
ZZ4	Sexuality	Grouped categories <ol style="list-style-type: none"> 1. Heterosexual 2. Homosexual/bisexual 3. Other
ZZ5	Indigenous status	Grouped categories <ol style="list-style-type: none"> 1. Indigenous 2. Not Indigenous
ZZ11	Industry	ANZSIC (grouped as described above)
ZZ12	Occupation	ANZCO (grouped as described above)
ZZ13a	Highest year of school completed	Original
Z5	Pregnancy status	Original

These variables are included on the CURF but require specific conditions of use to be met. All users accessing data must agree and sign the conditions of use form.