

6 AIHW validation checks

AIHW validation checks

The AIHW will apply an editing process to validate the data before loading it into a national database. It is assumed that jurisdictions will also perform validation checks and fix any errors that they can before the data is sent to the AIHW. The editing process will take place in two stages (in consultation with the data providers):

1. **Validity checks** are used to ensure that values entered for each data element are within a valid numeric range. For example, responses to the data element *Injecting drug use* should only be coded within the range of 1-4 or as 9. A response that does not fall within this range has to be an error. Therefore, range edits should identify incorrect and missing codes.
2. **Logic checks** are used to ensure internal consistency between responses within individual unit records. For example, when the response for *injecting drug use* = 4 (never injected), the response for *Method of use for principal drug of concern* cannot = 3 (injects).

Validity checks are performed first, so that the logic checks can be performed on valid data. A summary report on the findings from the validity and logic checks will be sent to each jurisdiction to allow consultation to resolve invalid/illogical data. Once validation issues have been resolved, revised frequency and cross tabulation tables on a selection of variables will be sent to each jurisdiction for approval to load the data into the national database. It is assumed that jurisdictions will have amended their own data sets with the agreed changes arising from the invalid/illogical data queries. It is also assumed that jurisdictions will have produced revised frequencies and cross-tabulation tables for checking against those supplied by the AIHW.

Table 4 contains a range of proposed validity checks to be applied to each State/Territory data set. It describes the range of values considered valid in the AODTS-NMDS as well as the treatment of 'not stated' or 'null' responses for each data element in the establishment-level and client-level collections.

Table 4: Proposed checks for valid dates and codes during preliminary loading

Data element	Validity check performed
Establishment-level	
Establishment identifier	Check that it is present for all unit records. Check for valid State identifier
Establishment type	Check for invalid code
Geographical location of establishment	Check against SLA table for invalid codes
Client-level	
Establishment identifier	Check for missing ID (must be present for all unit records)
Person identifier	Check for missing ID (must be present for all unit records)

Table 4 (continued): Proposed checks for valid dates and codes during preliminary loading

Data element	Validity check performed
Sex	Check for invalid code
Date of birth	Check for invalid date
Country of birth	Check against SACC table for invalid codes
Indigenous status	Check for invalid code
Preferred language	Check for invalid code
Client type – alcohol and other drug treatment services	Check for invalid code
Source of referral to AODT service	Check for invalid code
Date of commencement of treatment episode for alcohol and other drugs	Check for missing data (must be present for all unit records) and for invalid date
Date of cessation of treatment episode for alcohol and other drugs	Check for missing data (must be present for all unit records) and for invalid date
Reason for cessation of treatment episode for alcohol and other drugs	Check for invalid code
Number of service contacts within a treatment episode for alcohol and other drugs	Check for a numeric response that falls within an expected range (i.e. check against duration of episode established from commencement and cessation dates).
Treatment delivery setting for alcohol and other drugs	Check for missing data (must be present for all unit records) and for invalid code
Method of use for principal drug of concern	Check for invalid code
Injecting drug use	Check for invalid code
Principal drug of concern	Check for missing data (must be present for all unit records) and check against ASCDC table for invalid codes
Other drugs of concern	Check against ASCDC for invalid codes
Main treatment type for alcohol and other drugs	Check for missing data (must be present for all unit records) and for invalid code
Other treatment types for alcohol and other drugs	Check for invalid code

Table 5 contains a range of proposed logic checks to be applied to each individual record within a State/Territory data set after initial validity checks and preliminary loading. For these checks the output will identify records which do not pass the logic tests. These checks describe the types of data coding errors that are most likely to occur and will provide a guide for jurisdictions on the potential problem areas.

Table 5: Proposed logic checks for individual records

Data element	Logic check	Reasoning
Establishment identifier	<i>Establishment identifier</i> included in the Client file = the <i>Establishment identifier</i> in the Establishment file.	To ensure that all client record data can be linked to a valid establishment.
Geographical location of establishment	The first digit for <i>Geographical location of establishment</i> must = <i>State identifier</i> .	To ensure that the correct State/Territory identifier has been added to the SLA code.
Date of birth	Check if <i>Date of birth</i> is > than <i>Date of commencement of treatment episode for alcohol and other drugs</i> . Check if <i>Date of birth</i> before 1 January 1902, excluding 01011900 or after 01061993.	A client's <i>Date of birth</i> must be prior to their commencement of treatment. Check for outliers.
Country of birth and Preferred language	Check for abnormally large or unexpected frequencies in either data item, e.g. Preferred language = 00 (Afrikaans) where 00 is incorrectly used to code missing or not stated (98).	There is usually a connection between <i>Country of birth</i> and <i>Preferred language</i> .
Date of commencement of treatment episode for alcohol and other drugs	Check that <i>Date of commencement of treatment episode for alcohol and other drugs</i> is before 1 July 2003. <i>Date of commencement of treatment episode for alcohol and other drugs</i> must be equal to or prior to <i>Date of cessation of treatment episode for alcohol and other drugs</i> . Must also be after <i>Date of birth</i> .	Unit of measurement is all treatment episodes that closed in the period 1 July 2002 and 30 June 2003.
Date of cessation of treatment episode for alcohol and other drugs	Check that <i>Date of cessation of treatment episode for alcohol and other drugs</i> falls within 1 July 2002 and 30 June 2003. <i>Date of cessation of treatment episode for alcohol and other drugs</i> must be equal to or after <i>Date of commencement of treatment episode for alcohol and other drugs</i> .	Unit of measurement is all treatment episodes that closed in the period 1 July 2002 to 30 June 2003.
Reason for cessation of treatment episode for alcohol and other drugs	When <i>Reason for cessation of treatment episode for alcohol and other drugs</i> = 2, 3 or 4 check that next treatment episode for client reflects the correct change.	Provides a check on the validity of the trigger for a new treatment episode.
Treatment delivery setting for alcohol and other drugs	When <i>Treatment delivery setting for alcohol and other drugs</i> = 2 (residential treatment facility), check that <i>Number of service contacts within a treatment episode for alcohol and other drugs</i> = 0.	The number of service contacts is not collected for residential clients so for these clients the number should be set at 0.

(continued)

Table 5 (continued): Proposed logic checks for individual records

Data element	Logic check	Reasoning
Method of use for principal drug of concern	<p>Check if <i>Method of use for principal drug of concern</i> = 3 (injects), and <i>injecting drug use</i> = 4 (never injected).</p> <p>Check if <i>Method of use for principal drug of concern</i> is an appropriate code for example if <i>Principal drug of concern</i> = 2101 (alcohol) then <i>Method of use</i> should = 1 (ingest). See Appendix D for further information.</p>	<p>If a client states that they inject their <i>Principal drug of concern</i>, they cannot report that they have never injected.</p> <p>For example, if a client states that alcohol is their <i>Principal drug of concern</i>, it does not make sense if they report sniffing (powder) as their <i>Method of use for principal drug of concern</i>.</p>
Other drugs of concern	<p>Check if <i>Other drugs of concern</i> = <i>Principal drug of concern</i>.</p> <p>Check if any of the <i>Other drugs of concern</i> have the same code as each other.</p>	<p>This item complements <i>Principal drug of concern</i> and records any drugs apart from the principal drug.</p>
Main treatment type for alcohol and other drugs	<p>If <i>Main treatment type for alcohol and other drugs</i> = 5, 6 or 7 check that no <i>Other treatment type for alcohol and other drugs</i> is recorded.</p>	<p>If <i>Main treatment type</i> is support and case management only, information and education only, or assessment only then no other treatment should be reported.</p>
Other treatment type for alcohol and other drugs	<p>Check if <i>Other treatment type for alcohol and other drugs</i> = <i>Main treatment type for alcohol and other drugs</i>.</p> <p>Check if any of the <i>Other treatment types for alcohol and other drugs</i> have the same code as each other.</p>	<p><i>Other treatment type for alcohol and other drugs</i> is intended to only report any form of treatment that a client receives in addition to the <i>Main treatment type for alcohol and other drugs</i>.</p>

On completion of the proposed validity and logic checks, AIHW will produce frequency counts for the majority of variables in each State/Territory data set (see Table 6). Frequency tables will be used to check that data provided is consistent with the national standard and that frequency distributions are sensible. Some cross-tabulation counts will also be produced (see Table 7). AIHW produced frequency and cross tabulation counts will be compared with those provided by the States and Territories. The AIHW will consult with the relevant jurisdiction to resolve any differences.

Table 6: Proposed frequency tables for State and Territory data sets

Data element	Output labels
Establishment identifier	List of establishment identifiers
Establishment type	Public community centre Private community centre Public AODT centre Private AODT centre Not stated/inadequately described
Person identifier	Number of individual person identifiers (as a single aggregate number) contained in unit records
Sex	Male Female Indeterminate Not stated/inadequately described
Date of birth	Age groups (derived by subtracting each client's <i>Date of birth</i> from <i>Date of cessation of treatment episode for alcohol and other drugs</i>) < 20 20–29 30–39 40–49 50–59 60+
Country of birth	Frequency counts for all countries listed
Indigenous status	Aboriginal but not Torres Strait Islander Torres Strait Islander but not Aboriginal Aboriginal and Torres Strait Islander Not Aboriginal or Torres Strait Islander Not stated
Preferred language	Frequency counts for all languages listed (code range 00–98).
Client type – alcohol and other drug treatment	Own drug use Other's drug use Both own and other's drug use
Source of referral to AODT service	Frequency counts for all codes listed
Reason for cessation of treatment episode for alcohol and other drugs	Frequency counts for all codes listed

(continued)

Table 6 (continued): Proposed frequency tables for State and Territory data sets

Data element	Output labels
Number of service contacts within a treatment episode for alcohol and other drugs	Aggregate number of contacts reported and average number of contacts reported for treatment episodes (total no. of contacts/no. of treatment episodes). Note that treatment episodes in a residential setting should be excluded from any calculation.
Treatment delivery setting for alcohol and other drugs	Non-residential treatment facility Residential treatment facility Home Outreach setting Other
Method of use for principal drug of concern	Ingests Smokes Injects Sniffs (powder) Inhales (vapour) Other Not stated/inadequately described
Injecting drug use	Current injecting drug use Injecting drug use more than 3 months ago but less than 12 months ago Injecting drug use 12 months ago or more Never injected Not stated/inadequately described
Principal drug of concern	Alcohol (2101) Amphetamines (3100) Benzodiazepines (2400) Cannabis (3201) Cocaine (3903) Ecstasy (3405) Heroin (1202) Methadone (1305) Nicotine (3906) Other drugs
Other drugs of concern	Frequency counts (as above) for each of the first five other drugs of concern

(continued)

Table 6 (continued): Proposed frequency tables for State and Territory data sets

Data element	Output labels
Main treatment type for alcohol and other drugs	Withdrawal management (detoxification) Counselling Rehabilitation Pharmacotherapy Support and case management only Information and education only Assessment only Other
Other treatment type for alcohol and other drugs	Frequency counts for each of the first four other treatment types

Table 7: Proposed cross-tabulations for State and Territory data sets

Cross tabulation	Reasoning
Client type - alcohol and other drug treatment service by sex	This will check the male, female and total persons count by <i>Client type</i> to ensure that the variable has been correctly loaded and the coding is correct.
Indigenous status by sex	This will check the male, female and total persons count by <i>Indigenous status</i> to ensure that the variable has been correctly loaded and the coding is correct.
Principal drug of concern (selection) by sex Proposed drugs of concern: Alcohol (2101) Amphetamines (3100) Benzodiazepines (2400) Cannabis (3201) Cocaine (3903) Ecstasy (3405) Heroin (1202) Methadone (1305) Nicotine (3906)	This will check the male, female and total persons count by selections of <i>Principal drug of concern</i> to ensure that the variable has been correctly loaded and coding is correct.

Table 7 (continued): Proposed cross-tabulations for State and Territory data sets.

Cross tabulation	Reasoning
Age groups (derived by subtracting each client's Date of birth from Date of commencement of treatment episode) by sex and total persons <i>Proposed age groups:</i> < 20 20-29 30-39 40-49 50-59 60+	This will check the male, female and total persons count by derived age groups to ensure that <i>Date of birth</i> has been correctly loaded.
Treatment delivery setting for alcohol and other drugs by sex	This will check the male, female and total persons count by treatment delivery settings to ensure that <i>Treatment delivery setting for alcohol and other drugs</i> has been correctly loaded.
Main treatment type for alcohol and other drugs by Treatment delivery setting for alcohol and other drugs	This will check the treatment delivery setting count by main treatment types to ensure that <i>Main treatment type for alcohol and other drugs</i> has been correctly loaded.

Database sign-off

Before the AIHW collates the validated data into a national database, each jurisdiction will be required to 'sign-off' their data. The AIHW will provide each jurisdiction with frequency tables of their data set. If it is agreed that these tables are accurate, the jurisdiction will approve the AIHW to store the data into the national database and analyse it for the national report. This process keeps States and Territories fully informed of any changes or alterations that may be made to their data before it is used to produce the national report.

Each jurisdiction will also be given opportunities to view and comment on their data as presented in the national report before it is finalised.

Time-line for the validation process

The time-line for the validation process depends on the timely supply of the data from jurisdictions. Late supply may delay this process. The AIHW plans to send the 2002-03 validation tables to jurisdictions at the beginning of March 2004. Changes will be incorporated during March/April 2004. Tables for publication will be sent with the first draft of the national report for validation and approval.

Collection output

The AIHW is responsible for producing a comprehensive annual report on the AODTS-NMDS. The primary focus of the AIHW will be to produce national data, although where appropriate, State and Territory specific information may be produced. The AIHW may also consider ad hoc data requests (subject to confidentiality constraints and ethical clearance).

The annual report will be available in both hard copy and electronic form (PDF downloadable format) via the Institute's web site (www.aihw.gov.au).

Future data development

Development of the AODTS-NMDS will be ongoing and directed by the requirements of the IGCD and the States and Territories, in consultation with the AIHW and the Commonwealth. Development will include amending existing data elements and formulating new data elements when the need arises. Developing existing data elements will include refining data definitions and data domains, and modifying the directions in the 'guide for use' sections etc, as stakeholders identify problems. Potential data elements that increase the quantity and quality of the information collected by the AODTS-NMDS will continue to be developed.