

# 6 AIHW validation

## 6.1 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 (as specified in Tables 6 and 7) and fix any errors that they can before the data are 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 (see Table 6). 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 (see Table 7). 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 then logic checks. AIHW also performs an initial check for duplicate records. A summary report on the findings from the validity and logic checks will be sent to each jurisdiction to allow them to resolve invalid/illogical data.

Once validation issues have been resolved each jurisdiction will send AIHW:

1. revised data files;
2. revised frequency and cross tabulation tables for checking against AIHW frequencies and cross-tabulations;
3. the validation report sent by the AIHW with changes made as a result of queries documented as **tracked changes** by the jurisdiction.

AIHW will then check the revised frequency and cross tabulation tables and the changes that have been made by the jurisdiction. When correct, AIHW will request that the jurisdiction signs off its data for loading to the national database.

Table 6 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 6: Proposed checks for valid dates and codes during preliminary loading**

Data element	Validity check performed
<b>Establishment-level</b>	
Establishment identifier	Check that it is present for all unit records. Check for valid State identifier and valid layout (see table 3).
Geographical location of service delivery outlet	Check against SLA table for invalid codes. Use the ASGC publication relevant to the data collection year i.e. 2003.
<b>Client-level</b>	
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) Use alpha, numeric or alphanumeric characters only

(continued)

**Table 6 (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, including future dates
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
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 7 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. 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 7: Proposed logic checks for individual records**

Data elements	Logic check	Reasoning
Establishment identifier	<i>Establishment identifier</i> included in the Client file = the <i>Establishment identifier</i> in the Establishment file and there should be the same number of <i>Establishment identifiers</i> in each file (allowing for repetition of <i>Establishment identifiers</i> in client file).	To ensure that all client record data can be linked to a valid establishment.
Geographical location of service delivery outlet/ establishment identifier	The first digit for <i>Geographical location of service delivery outlet</i> must = <i>State identifier</i> in the <i>Establishment identifier</i> .	To ensure that the correct state/territory identifier has been added to the SLA code.
Date of birth and Date of commencement	Check if <i>Date of birth</i> is > than <i>Date of commencement of treatment episode for alcohol and other drugs</i> .	A client's <i>Date of birth</i> must be prior to their commencement of treatment.

(continued)

**Table 7 (continued): Proposed logic checks for individual records**

<b>Data elements</b>	<b>Logic check</b>	<b>Reasoning</b>
Date of birth and Date of cessation	Check if <i>Date of birth</i> before 1 January 1902, excluding 01011900, or is a date that provides a client age of less than 10 years when age is based on <i>Date of cessation of treatment episode</i> .	Check for outliers.
Country of birth and Preferred language	Check for abnormally large or unexpected frequencies in either data item, e.g. <i>Preferred language</i> = 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>
Country of birth and Indigenous status	Check for records where <i>Indigenous status</i> = 4 and <i>Preferred language</i> = 2, 5, 7, 12, 41, 42, 54, 55, 76, 82, 83 or 85 (codes for Indigenous languages).	There is usually a connection between <i>Country of birth</i> and <i>Indigenous status</i>
Indigenous status and Country of birth	Check for records where <i>Indigenous status</i> = 1, 2 or 3 and <i>COB</i> is not Australia or External Territories.	There is usually a connection between these codes.
Date of commencement and Date of cessation and Date of birth	<i>Date of commencement of treatment episode for alcohol and other drugs</i> must be equal 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 2003 and 30 June 2004.
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 2003 and 30 June 2004.	Unit of measurement is all treatment episodes that closed in the period 1 July 2003 to 30 June 2004.
Reason for cessation of treatment episode for alcohol and other drugs	When Reason for cessation of treatment episode = 2 (Main treatment type) check that the next treatment episode for client reflects the correct change.  When Reason for cessation of treatment episode = 3 (Principal drug) check that the next treatment episode for client reflects the correct change.  When Reason for cessation of treatment episode = 4 (Treatment delivery setting) check that the next treatment episode for client reflects the correct change.	Provides a check on the validity of the trigger for a new treatment episode.
Method of use for principal drug of concern and Injecting drug use	Check if Method of use for principal drug of concern = 3 (injects), and Injecting drug use = 4 (never injected).	If a client states that they inject their Principal drug of concern, they cannot report that they have never injected.
Method of use and Principal drug of concern	Check if Method of use for principal drug of concern is an appropriate code for example if Principal drug of concern = 2101 (alcohol) then Method of use should = 1 (ingest).	For example, if a client states that alcohol is their Principal drug of concern, it does not make sense if they report sniffing (powder) as their Method of use for principal drug of concern.
Other drugs of concern and Principal drug of concern	Check if <i>Other drugs of concern</i> = <i>Principal drug of concern</i> .  Check if any of the <i>Other drugs of concern</i> have the same code as each other.	This item complements <i>Principal drug of concern</i> and records any drugs apart from the <i>Principal drug</i> .
Main treatment type and Client type	If Client type = 2 check that Main treatment type for alcohol and other drugs does not = 1, 3 or 4.	These are not appropriate codes for a client receiving treatment/ advice for the drug use of others.
Other treatment type for alcohol and other drugs and Main treatment type	Check if Other treatment type for alcohol and other drugs = Main treatment type for alcohol and other drugs.	This item complements <i>Main treatment type</i> and records any treatments apart from the <i>Main treatment</i> .

(continued)

**Table 7 (continued): Proposed logic checks for individual records**

Data elements	Logic check	Reasoning
Other treatment type	Check if any of the Other treatment types for alcohol and other drugs have the same code as each other.	<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> .
Other treatment type and Client type	If Client type = 2 check that Other treatment type for alcohol and other drugs does not = 1, 3 or 4.	These are not appropriate codes for a client receiving treatment/advice for the drug use of others.

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 8). 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 9). AIHW-produced frequency and cross tabulation counts will be compared with those provided by the states, territories and Commonwealth. The AIHW will consult with the relevant jurisdiction to resolve any differences.

**Table 8: Proposed frequency tables for state and territory data sets**

Data element	Output labels
Establishment identifier	List of establishment identifiers
Sex	Male Female Not stated/inadequately described
Date of birth	Age groups (derived by subtracting each client's Date of birth from Date of <b>cessation</b> of treatment episode for alcohol and other drugs) (see also AIHW SAS program under cross-tabulations for Age by sex on p. 39). <b>Note:</b> clients aged under 10 years will not be reported on in 2002–03. 10–19 20–29 30–39 40–49 50–59 60+ missing (include clients with default date of 01011900)
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
Client type—alcohol and other drug treatment	Own drug use Other's drug use

(continued)

**Table 8 (continued): Proposed frequency tables for state and territory data sets**

<b>Data element</b>	<b>Output labels</b>
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
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	Last injected within the previous three months Last injected more than three months ago but within the previous twelve months Last injected more than twelve months ago Never injected Not stated/inadequately described
Principal drug of concern <i>Please use agreed short list</i>	Alcohol (2101) Amphetamines (3100) Benzodiazepines (2400) Cannabis (3201) Cocaine (3903) Ecstasy (3405) Heroin (1202) Methadone (1305) Nicotine (3906) Other drugs (balance of ASCDC codes)
Other drug of concern <i>Please use agreed short list</i>	Frequency counts (as above) for each of the five other drugs of concern as separate counts.
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 four other treatment types as separate counts.

**Table 9: 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.  <i>Please use agreed short list</i>
Age groups (derived by subtracting each client's Date of birth from Date of <b>cessation</b> of treatment episode) by sex and total persons  Proposed age groups: 10–19 20–29 30–39 40–49 50–59 60+ missing	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.  Age groups format (SAS program) <pre> data data01_freq_date_of_birth ; set access01.episod01; keep establis person_i date_of_birth sex date_of_cessation state age; age = floor ((date_of_cessation - date_of_birth)/365.25); libname formats ' '; proc format library = work.formats ; value agrpfmt 0 - 9.9999 = '0-9' (as a check only not included in table) 10 - 19.9999 = '10-19' 20 - 29.9999 = '20-29' 30 - 39.9999 = '30-39' 40 - 49.9999 = '40-49' 50 - 59.9999 = '50-59' 60 - 99.9999 = '60+' 100 - high = 'missing';           </pre>
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 <i>Treatment delivery setting</i> count by <i>Main treatment types</i> 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 states and territories make their own changes or alterations to their data on the basis of the Validation report sent to them by the AIHW and resend their final revised data file to the AIHW. The AIHW will check the revised frequency and cross-tabulation tables provided by the jurisdiction for 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. The data set held by each jurisdiction will match the data set held by AIHW.

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 hinges on the timely supply of the data from jurisdictions. Late supply will delay this process. The AIHW plans to send the 2003–04 validation reports to jurisdictions progressively within 2–3 weeks of receipt of data. On receipt and verification of the revised data set and frequency and cross-tabulation tables from all jurisdictions AIHW will commence analysis of the data for the national report. Tables for publication will be sent with the first draft of the national report for validation and approval.

**Table 10: National timetable for transfer, validation and reporting of 2003–04 data**

<b>Year 2004</b>		
<b>All dates to be confirmed at April WG meeting</b>		
August	Jurisdictions	Jurisdictions to commence process of receiving and cleaning 2003–04 data from agencies
October	AIHW	Formal request for the 2003–04 data
Oct-Dec 31	Jurisdictions	Transfer of clean data (2003–04) to the AIHW with file specifications and frequency and cross-tabulation tables
<b>Year 2005</b>		
January	AIHW	Draft of 2005–06 Guidelines circulated to jurisdictions for feedback
	Jurisdictions	Send comments to AIHW on 2005–06 Guidelines
January	AIHW	Commencement of validation process
January to March	AIHW	Validation report and data queries sent to each jurisdiction
January to March	Jurisdictions	Respond to validation report and data queries. Send revised data file and revised frequency and cross-tabulation tables.
March to April	AIHW & jurisdictions	Final data check and verification. Sign-off provided by jurisdictions for final loading of data
February	AIHW	2005–06 Guidelines placed on AIHW web site
early April	AIHW	Commence analysis for reporting on 2003–04 data
early May	AIHW	First draft of '2003–04 national report' circulated internally
late May	AIHW	Draft of '2003–04 national report' circulated to jurisdictions for comment
August	AIHW	Final draft of 2003–04 national report circulated for comments and final editing
October	AIHW	Release of 2003–04 national report

## **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 format) via the Institute's web site ([www.aihw.gov.au](http://www.aihw.gov.au)).

## **Future data development**

Development of the AODTS-NMDS will be directed by the requirements of the IGCD and the states and territories, in consultation with the AIHW and the Commonwealth. For the 2003-04 collection the emphasis will be on consolidating the existing AODTS NMDS. Enhancements to existing data elements may include refining data definitions and data domains, and modifying the directions in the 'guide for use' sections etc, as stakeholders identify problems. Future development will include amending existing data elements and formulating new data elements when the need arises. Potential data elements that increase the quantity and quality of the information collected by the AODTS-NMDS will continue to be developed as the need arises.

## ATTACHMENT A

### Documentation of 2003–04 alcohol and other drug treatment services data

PLEASE COMPLETE THIS FORM AND RETURN IT TO THE AIHW WITH YOUR 2003–04 DATA.

#### Instructions

To assist the AIHW in the processing of alcohol and other drug treatment services data, it is requested that each jurisdiction complete the following documentation. Please answer all questions and supply the relevant information where requested. Of particular importance is information for variables that do not comply with the national standard. If the codes or descriptions used by agencies in your state/territory differ from the national standard, please document these differences and provide information necessary for the translation or manipulation of the data to achieve national standards.

Also, please indicate if you have not provided data for any of the data elements requested. Please insert the state/territory name in the box below:

#### Specific details about the data supplied

##### Format of data

Please indicate the format in which you have provided the data to the Institute.

- Flat file – comma separated variable length records (CSV)\*
- Flat file – fixed length records with data items identified by position
- SAS file – SAS transport file
- SAS file – PC SAS file
- MS Access file
- MS Excel file

*Please note that all data must be in ASCII format and unformatted (i.e. stripped of all applied formats).*

\* This is the AIHW preferred format

**Any other comments?**

## Establishment data file

### List of agencies/establishments included in 2003–04 collection

It is important for the AIHW to know whether the list of agencies contributing data for the 2003–04 collection has changed from the 2002–03 collection.

The AIHW would like each jurisdiction to insert or attach a list of establishments/agencies (including establishment number and the name of the agency) that have contributed data for the 2003–04 collection.

Please indicate any agencies that are new to the collection (i.e. were not included in the 2002–03 collection) or have changed their establishment number or agency name from the previous year.

It is very important that the AIHW is informed of any agencies within scope for the collection that have not provided any data for the 2003–04 period, or that were not able to provide data for the entire collection period.

**[Insert the list here or attach at end of document]**

**Any other comments?**

### File positions of variables

File position	Variable name
1	Establishment identifier
2	Geographical location of service delivery outlet

1. Are these positions correct? Yes    
No

If not correct, please provide details. Also, if a tab-delimited file has been used, please indicate the positions and lengths of the variables in the file.

## National standards

2. Has the correct national standard been used for all establishment-level data elements?

Yes [ ]

No [ ]

If not, please provide details.

## Specific questions

3a. Are *Establishment identifiers* stable, unique identifiers for your state/territory?

If not, how are records uniquely identified by your state/territory?

3b. Do you have unique *Establishment identifiers* for each *Service delivery outlet*? *Service delivery outlet* refers to the **site** from which an organisation, or sub-unit of an organisation, delivers a health/community service. That is, agency sites that conduct treatment as distinct from administrative centres. Accordingly, each site should have a unique *Establishment identifier*. The Establishment number comprises the last 5 digits of the *Establishment identifier* and provides space for the *Service delivery outlet* number.

3c. Do you have any comments on *Service delivery outlet*?

4. Please ensure that the code used for your State/territory identifier, corresponds to the correct code as shown below.

NMDS Code	NMDS Description
1	New South Wales
2	Victoria
3	Queensland
4	South Australia
5	Western Australia
6	Tasmania
7	Northern Territory
8	Australian Capital Territory
9	Other territories



### File positions of variables (continued)

File position	Variable name
9	Source of referral to alcohol & other drug treatment service
10	Date of commencement of treatment episode
11	Date of cessation of treatment episode
12	Reason for cessation of treatment episode
14	Treatment delivery setting for alcohol and other drugs
15	Method of use for principal drug of concern
16	Injecting drug use
17	Principal drug of concern
18	Other drug of concern (a)
19	Other drug of concern (b)
20	Other drug of concern (c)
21	Other drug of concern (d)
22	Other drug of concern (e)
23	Main treatment type for alcohol and other drugs
24	Other treatment type (a)
25	Other treatment type (b)
26	Other treatment type (c)
27	Other treatment type (d)

7. Are these positions correct? Yes  [ ]  
 No  [ ]

If not, please provide details in the following section, including details of any Not applicable codes used, what code has been used and for which data items.

### National standards

It is very important that the AIHW is informed of all cases where national standards have not been used or where mapping to the national standard has occurred.

8. Has the correct national standard been used for all client-level data elements?  
 Yes  [ ]  
 No  [ ]

If not, please provide details, including details of any Not applicable codes used, what code has been used and for which data items.

### Specific questions

9. Is the *Person identifier* maintained for individuals when they re-register for treatment? If so, does this apply only within individual agencies or throughout the state/territory? Will the identifiers continue to be used in following years? (That is, will it be possible to identify new episodes recorded for the same clients as registered in 2003-04?).

10. When *Date of birth* was not available, has an estimate date been provided? If so, please provide detail about how estimates are derived.
11. Has *Country of birth* been coded according to the Standard Australian Classification of Countries (SACC) ABS. cat. no. 1269.0 (1998)? If not, please document what standard has been used? If a select short list of most common countries has been used, please provide this list and the respective coding used.
12. Please comment on the quality of data reported for *Indigenous status*. Is the quality considered acceptable or in need of improvement? Please describe any known limitations on the quality of the data.
13. If, for any reason, any of the following variables have not been coded in accordance with the 2002–03 Guidelines standard please indicate which variable (s) and what alternative coding method has been used.
  - *Preferred language*
  - *Client type*
  - *Source of referral to alcohol and other drug treatment service*
  - *Reason for cessation*
  - *Method of use*
  - *Injecting drug use*
  - *Principal drug of concern*
  - *Main treatment type*
14. As *Other drug of concern* is a multiple response variable, it is important that you indicate the maximum number of responses possible for each client. That is, how many other drugs can be recorded for each client (e.g. 1–5 or more than 5)? If a CSV file has been sent, you must indicate the number of commas (positions) that have been allocated for this variable on each unit record. Note: code 0003 (no *Other drug of concern*) should only be used for the first *Other drug of concern*.
15. As *Other treatment type for alcohol and other drugs* is a multiple response variable, it is important that you indicate the maximum number of responses possible for each client. That is, how many other treatment types can be recorded for each client (e.g. 1–4, more than 4?) and how many have been reported to AIHW (if number different to

number recorded)? If a CSV file has been sent, you must indicate the number of commas (positions) that have been allocated for this variable on each unit record.

16. Does a change in the *Treatment delivery setting* act as a trigger for a new treatment episode?
17. Does a change in the *Principal drug of concern* act as a trigger for a new treatment episode?
18. Does a change in the *Main treatment type* act as a trigger for a new treatment episode?
19. When *Client type* = Other's drug use (code 2) have the following data items been coded to Not Stated: *Method of use for principal drug of concern* (code 9), *Injecting drug use* (code 9), *Principal drug of concern* (code 0001), and *Other drugs of concern* (code 0001)?

### **Additional comments?**

Please add any other comments on data availability, quality and/or scope that are necessary for appropriate loading and analysis of these data.