

5 Collection procedures and data quality

Chapters 5 and 6 provide information on the data collection and transfer process for the 2003–04 collection and include data quality and validation checks. The information contained in these chapters is to be used by jurisdictions to prepare appropriate edit checks for the 2003–04 collection and for the cleaning of the 2003–04 data prior to transmission. It should also be used by jurisdictions to inform their agencies of the type and use of appropriate codes when collecting and collating the 2003–04 AODTS NMDS data. Previously a separate document *Data Transfer and Validation Procedures for Alcohol and Other Drug Treatment Services Data* was sent out in September each year to provide guidance on the validation and transfer process. As it was thought confusing to have two separate documents, the current year's Guidelines publication amalgamates the content of the two documents and should be used as the sole authority for the data transfer and validation process.

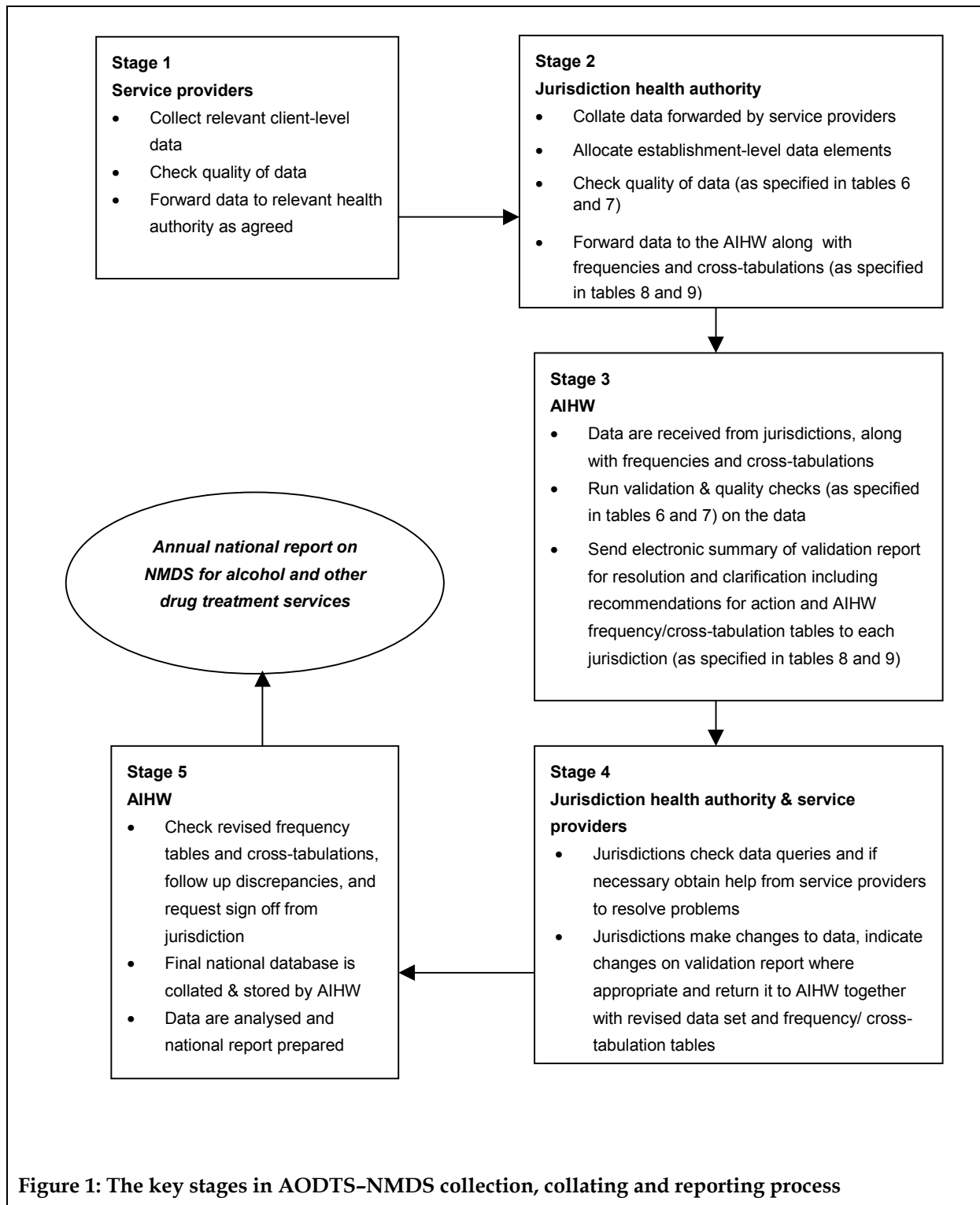
5.1 Collation of the national data set

The collation of a national data set involves five distinct stages (see Figure 1).

1. The first stage is the collection of the agreed data elements by service providers for each client who is eligible for inclusion in the collection. Service providers then forward their collected information to the designated health authority for collation. This process will differ across jurisdictions, as service providers in some states/territories are required to forward their data to an area or region coordinator, whereas in other states the data are forwarded directly to the central authority.
2. The second stage involves the designated health authority collating the data that were forwarded by the service providers. At this stage the data should also undergo a rigorous validation process to ensure the quality of the information using the validation checks in Tables 4 and 5. Health authorities are required to allocate establishment-level data elements. The collated unit record data is then forwarded to the AIHW.
3. At stage three the AIHW receives the collated state/territory data for validation. When finished validating the data, the AIHW sends an electronic summary validation report to each state/territory health authority (which includes all queries and identified problems with their data relating to checks specified in tables 4 and 5) for resolution and clarification. The report includes an initial set of frequency and cross-tabulation tables (as specified in Tables 6 and 7).
4. At stage four, state and territory health authorities assess which of the changes specified in the summary validation report need to be made to the data and then make those changes. Details of the changes made are to be added to the appropriate section of the summary validation report and the report returned electronically to AIHW together with an updated data file and revised frequency and cross-tabulation tables for final checking.
5. AIHW checks that the changes have been made and the revised frequency and cross-tabulation tables are correct. State and territory health authorities can then sign off their data set (i.e send an email to AIHW authorising the loading of their data to the national database). The jurisdictional data will then be loaded to the national database where all data are stored by the AIHW ready for analysis and reporting.

Note that:

- no data are to be directly submitted by service providers to the AIHW;
- the information transferred from service providers to health authorities and then to the AIHW does not include client names, only a person identifier code that is generated by the service provider.



5.2 Data transfer

Service providers to health authorities

Protocols for the transfer of data from alcohol and other drug treatment agencies to their jurisdictional health authority vary between states and territories. Each health authority responsible for the AODTS–NMDS collection will contact service providers within scope for the collection to inform them of the required format and timing of the data transfer.

Health authorities to AIHW

The NMDS data will need to be forwarded to the AIHW annually by each jurisdiction. The data requested will be for a financial year reference period (1 July to 30 June). Data for the period 1 July 2003 to 30 June 2004 will be requested by the AIHW early in the 2004–05 financial year (October 2004). It is expected that state and territory health authorities will supply this data to the AIHW by 31 December 2004. The results of the analysis of these data, at both the national and state/territory levels, will be reported during 2005.

Transfer method

The preferred transfer method is by email attachment. AIHW will also accept data on floppy disk. Note that floppy disks can only hold 1.4 Mb of data, and the AIHW can only accept files by email that are less than 4 Mb. Files sent by email or floppy disk should be compressed, preferably with WinZip, and password protected. Jurisdictions interested in sending the file via email are requested to contact the AIHW before sending the file to advise the AIHW of the password.

If a file does not fit on a floppy disk or is too large for an email attachment, the next preferred option is a CD-ROM.

File format

When jurisdictions are satisfied that their data are clean, and that all practical follow-up has been completed, unformatted data should be forwarded to the AIHW contact in the following form: **Comma Separated Variable (CSV) ASCII text records.**

For example, a single client unit record will look like the following:

```
12A00101, PID99, 1, 05061977, 1101, 4, 19, 1, 01, 02092001, 03122001, 07, 02, 1, 3201, 0003, , , 2, 4, 2, 8, , , ,
```

The following file types can be accepted by AIHW:

- Microsoft Excel file
- Microsoft Access file
- Unformatted SAS file in transport mode indicating the appropriate platform (e.g. Unix, NT, MVS) and the SAS version used.

File content

There should be two files for each jurisdiction:

- establishment-level file (statistical unit = alcohol and other drug treatment agency/organisation)
- client-level file (statistical unit = closed treatment episode).

Please ensure column descriptors are included for both files.

Accompanying information

When transferring data to the AIHW, each jurisdiction should include the following documentation:

1. **Summary frequency and cross-tabulation tables**, which are used by the AIHW to verify information when compiling the national data set; and
2. **Attachment A**. This attachment is designed to obtain a description of the file and to identify variables that do not conform to the standard definitions and any translation or manipulation of the data necessary to achieve national standards. This information will assist the AIHW to correctly load and interpret the data.

Mandatory data items

The following data items are mandatory data items. Where information is not available to code these data items the record should be excluded.

- Establishment identifier
- Person identifier
- Client type
- Principal drug of concern
- Main treatment type
- Treatment delivery setting

File specification

The following tables specify the order in which the data items should be provided to the AIHW and the range of valid responses.

Table 4: Specifications for data transfer to AIHW of establishment-level data

Item no.	Item name	Description	Valid response
1	Establishment identifier	Alphanumeric item. Consists of a unique identifier for each state/territory (<i>State identifier</i>), an establishment sector code (<i>Establishment sector</i>), an alphanumeric, alpha or numeric code for the location of the service (<i>Region code</i>), and a unique service number within the state/territory which should include a service delivery outlet code where applicable (<i>Establishment number</i>).	Maximum 9 characters (minimum of 8) NNANNNNN where N = <i>State identifier</i> (valid range 1-9), N = <i>Establishment sector</i> (valid range 1-2), AA = <i>Region code</i> (jurisdiction specific), NNNNN = <i>Establishment number</i> (jurisdiction specific and including a code for service delivery outlet). Note that a 'not stated', 'null' or 'missing' response is not permitted. Where no Establishment identifier has been assigned, the establishment record should be excluded.
2	Geographical location of service delivery outlet	Numeric code to indicate the statistical local area (SLA) of the service delivery outlet within the reporting state or territory.	Five-digit valid code as defined in the <i>Australian Standard Geographical Classification</i> (ASGC 2003), ABS Cat. No. 1216.0.

Table 5: Specifications for data transfer to AIHW of client-level data

Item no.	Item name	Description	Valid response
1	Establishment identifier	As before.	As before.
2	Person identifier	Alpha, numeric or alphanumeric item. A client's unique identifier within an agency.	Alpha, numeric or alphanumeric code (agency determined). Non alphanumeric characters are not valid e.g. / # etc. Note that a 'not stated', 'null', 'missing' or 'coding error' response is not permitted. Where no person identifier has been assigned, the client record should be excluded.
3	Sex	Numeric code to indicate the sex of a client.	1 (male), 2 (female) or 9 (not stated/ inadequately described). Use code 9 as the default value for invalid codes and missing values.
4	Date of birth	Numeric item stating a client's date of birth.	DDMMYYYY format, right justified, zero-filled (e.g. 3 March 1965 would be 03031965). When an estimate is required use 0101 with a valid year with a flag to indicate that it is an estimate . Note: year must always be four-digit format (YYYY). Use 01011900 as default date for invalid and missing dates.
5	Country of birth	A numeric code to indicate the country in which a client was born.	Four-digit valid code as defined in the <i>Standard Australian Classification of Countries</i> (SACC), ABS Cat. No. 1269.0 (1998). Includes 0000 (inadequately described), 0001 (at sea), 0002 (not elsewhere classified) and 0003 (not stated). Use 0000 as default value for invalid codes and 0003 for missing data. Codes 0004 and 0005 are not valid responses. They apply to economic statistics only.
6	Indigenous status	A numeric code indicating whether or not a client is of Aboriginal and/or Torres Strait Islander origin.	1–4 and 9 (not stated). Code 9 should only be used when the question was refused or not able to be asked.
7	Preferred language	A numeric code indicating the language most preferred by a client for communication.	Two-digit valid code 00-86 and 95 (Other languages nfd), 96 (Inadequately described), 97 (Non-verbal, so described including sign languages) and 98 (Not stated). Right justified. Use 96 as default value for invalid codes and 98 for missing values.
8	Client type— alcohol and other drug treatment services	A numeric code indicating the status of a person in terms of whether contact with the agency concerns their own alcohol and/or other drug use or that of another person.	1 (own drug use) 2 (other's drug use). Note that a 'not stated', 'null', 'missing', or 'not known' response is not permitted. Where client is attending for both his/her own drug use and the drug use of another person code 1 should be used.
9	Source of referral to alcohol and other drug treatment service	A numeric code indicating the source from which a client was transferred or referred to an agency.	1–10, 98 (other) and 99 (not stated/ inadequately described), right justified. Use code 99 as the default value for invalid codes and missing values (see also pp. 15–16).
10	Date of commencement of treatment episode for alcohol and other drugs	Numeric item stating the date on which a client's treatment episode commenced.	DDMMYYYY format, right justified, zero-filled (e.g. 3 March 2003 would be 03032003). Note that a 'not stated', 'null', 'missing', 'not known' or an invalid date response is not permitted. Where an accurate date of commencement is not available, the client record should be excluded.

(continued)

Table 5 (continued): Specifications for data transfer to AIHW of client-level data

Item no.	Item name	Description	Valid response
11	Date of cessation of treatment episode for alcohol and other drugs	Numeric item stating the date on which a client's treatment episode ceased.	DDMMYYYY format, right justified, zero-filled (e.g. 3 March 2003 would be 03032003). Note that a 'not stated', 'null', 'missing', 'not known' or an invalid date response is not permitted. Where an accurate date of cessation is not available, the client record should be excluded.
12	Reason for cessation of treatment episode for alcohol and other drugs	Numeric item stating the reason a client's treatment episode ceased.	1–13, 98 (Other) and 99 (Not stated/inadequately described).
13	Treatment delivery setting for alcohol and other drugs	Numeric item stating the type of treatment delivery setting.	1–4 and 8 (Other) Note that a 'not stated', 'null', 'missing' or 'not known' response is not permitted. Where no treatment delivery setting has been assigned, the client record should be excluded.
14	Method of use for principal drug of concern	A numeric code indicating a client's usual method of administering principal drug of concern, as stated by the client.	1–6 and 9 (not stated/inadequately described). Use code 9 as the default value for invalid codes and missing values, and when Client type = 2.
15	Injecting drug use	A numeric code indicating a client's use of injection as a method of administering drugs.	1–4 and 9 (not stated/inadequately described). Use code 9 as the default value for invalid codes and missing values and when Client type =2.
16	Principal drug of concern	A numeric code indicating the drug that has led a client to seek treatment or advice from an agency. As stated by the client. If client has been referred into treatment and does not nominate a principal drug then use the drug involved in the client's referral.	Four-digit valid code as defined in the <i>Australian Standard Classification of Drugs of Concern</i> , ABS Cat. No. 1248.0 (2000). Note that a 'not stated', 'null', 'missing' or 'not known' response is not permitted. Where no principal drug has been assigned, the client record should be excluded. Use code 0000 for Inadequately described and code 0005 for Opioid analgesics nfd Use code 0001 when Client type = 2.
17a	Other drugs of concern (1st)	A numeric code indicating any drugs, in addition to the principal drug of concern, which a client perceives as a concern.	Four-digit valid code from <i>Australian Standard Classification of Drugs of Concern</i> , ABS Cat. No. 1248.0 (2000) or 0001 (not stated) and 0003 (none/no other drugs of concern). Use code 0001 as the default value for invalid codes and when Client type = 2. Code 0003 can only be used for the 1st Other drug of concern (see example below under 'ensure correct data loading'). Up to 5 other drugs of concern can be reported see items 14a–14e. Do not send more than 5 Other drugs of concern. To ensure correct data loading, 5 commas must be inserted even if blanks are present. (E.g. if there are no other drugs of concern it would be 0003, , , , if 2 other drugs of concern it would be NNNN, NNNN, , ,).
17b	Other drugs of concern (2nd)	A numeric code indicating the second other drug of concern.	Four-digit valid code as above. If blank response then insert a comma (,) without brackets.

(continued)

Table 5 (continued): Specifications for data transfer to AIHW of client-level data

Item no.	Item name	Description	Valid response
17c	Other drugs of concern (3rd)	A numeric code indicating the third other drug of concern.	Four-digit valid code as above. If blank response then insert a comma (,) without brackets.
17d	Other drugs of concern (4th)	A numeric code indicating the fourth other drug of concern.	Four-digit valid code as above. If blank response then insert a comma (,) without brackets.
17e	Other drugs of concern (5th)	A numeric code indicating the fifth other drug of concern.	Four-digit valid code as above. If blank response then insert a comma (,) without brackets.
18	Main treatment type for alcohol and other drugs	A numeric item stating the main treatment type.	1–8. Each treatment episode will only have one main treatment type. Note that a 'not stated', 'null', 'missing' or 'not known' response is not permitted. Where no main treatment type has been assigned, the client record should be excluded.
19a	Other treatment type for alcohol and other drugs (1st)	A numeric code indicating any treatment that a client is receiving in addition to the Main treatment type.	1–5 Up to 4 other treatment types can be reported, see Item nos. 20a–20d. To ensure correct data loading, 4 commas must be inserted, even if blank responses are present (e.g. if no other treatment types it would be ,,,, If 2 other treatment types it would be N, N,,).
19b	Other treatment type for alcohol and other drugs (2nd)	A numeric code indicating the second Other treatment type.	1–5. If a blank response then insert a comma (,) without brackets
19c	Other treatment type for alcohol and other drugs (3rd)	A numeric code indicating the third Other treatment type.	1–5. If a blank response then insert a comma (,) without brackets
19d	Other treatment type for alcohol and other drugs (4th)	A numeric code indicating the fourth Other treatment type.	1–5. If a blank response then insert a comma (,) without brackets

AIHW contacts for further information on file transfer

Ms Gail Weaving

Functioning and Disability Unit

Phone: (02) 6244 1050

Email: gail.weaving@aihw.gov.au

Ms Elena Ougrinovski

Data and Information Technology Unit

Phone: (02) 6244 1147

Email: elena.ougrinovski@aihw.gov.au

5.3 Data quality

Data collections require ongoing attention to quality. There is a need to attend to how questions are asked and information obtained, data entry, the handling of missing and erroneous information, edit checking and there is a need to follow-up with data providers to ensure the highest quality data possible.

To ensure that the AIHW is supplied with a useable national data set, it is essential that jurisdictions clean (edit) the data they receive from service providers before they transfer it to the AIHW. The quality of the NMDS data will also be enhanced if service providers check the quality of their data before sending it to their jurisdictional health authority. In collating the data into a national database, the AIHW also follows a formal validation process to maximise data quality (see Chapter 6).

General checks that should be conducted

Service providers and jurisdictions should perform the following quality checks before the data are sent to the AIHW.

- **Missing agencies:** Jurisdictions should ensure that all agencies within scope of the collection have sent data for the entire collection period.
- **Missing data:** Jurisdictions should investigate missing data to ensure that agencies are reporting all AODTS-NMDS data items. A reasonable attempt should be made to resolve missing data issues, at both an agency level and at the unit record level.
- **Incorrect codes:** Jurisdictions should ensure that agencies use the correct codes for all data items. This may involve mapping codes at the state/territory office before sending data to AIHW. Coding errors that cannot be corrected should be coded to the appropriate default value (e.g. inadequately described).
- **Duplicate records:** Jurisdictions should check for duplicate unit records. When records are identified as possible duplicates, the agency should be consulted to ensure that unit records have not been mistakenly submitted on more than one occasion. **The following data items are used by AIHW to check for duplicates:**
 - establishment identifier
 - person identifier
 - state
 - date of birth
 - date of commencement of treatment
 - date of cessation of treatment
 - principal drug of concern
 - main treatment type
 - treatment delivery setting
- **Reporting period:** The cessation dates of treatment episodes should be checked to ensure that only treatment episodes that closed within the valid reporting period (1 July 2002 to 30 June 2003) are included in the 2002–03 collection.
- **Data inclusion:** Jurisdictions should ensure that data not within scope of the AODTS-NMDS are excluded from the collated data set sent to the AIHW (e.g. methadone treatment where there are no main or other treatment types).

- **Establishment identifiers:** Jurisdictions should ensure that establishment identifiers used on the establishment data file are the same as those used on the client data file and that there are the same number of establishments on each file.
- **Geographical location of service delivery outlet:** Jurisdictions should ensure that all geographic location codes begin with a valid state or territory identification number, and are a valid SLA for the period in question. i.e. 2003-04.
- **Client type:** Jurisdictions should ensure that for clients who attend treatment because of another person's drug use (**client type = 2**), the following data elements are 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)
 - *Other drugs of concern* (code 0001)