

Appendix 4: Using the statistical linkage key

This appendix provides details of linkage processes, including validation of the linkage key, results of linkage and the rules used to allocate responses that are discrepant between linked records. It relates to material presented in Section 2.4 and in Chapter 4.

Methods for the linkage of records

The process for linking records is as follows:

- Two, three or more records that had fully valid linkage keys that completely matched were regarded as referring to the same consumer.
- Records that were missing sex were tested for matches with all other records using the linkage key without sex. This resulted in 14 matches (12 pairs and 1 triple), and for each match the records were from the same State. These matching records were also regarded as referring to the same consumer. For each match at least one case had sex validly recorded for the consumer.
- Records that were missing any part of the linkage key other than sex were not matched and were given a unique key for all future analyses.
- The letter part of the linkage key was examined to check whether any unlikely or possibly false sequences (such as 'ABCDE') or repetitions (such as 'AAAA-') appeared at a higher frequency than might be expected. The only such apparent sequence was 'XXXXX', which was obviously false as it occurred 4 times for only one service provider. For these records the linkage key was treated as invalid.
- For the date part of the key, the frequency distribution of days and months was examined for any unexpected patterns. As in 1999, it was apparent that 1 January was more common as a birth date than was expected. The birth date of some Aboriginal and Torres Strait Islander people may be legitimately recorded as 1 January when their precise birth date is not known. This would explain a proportion of these dates, as nearly 10% were for Aboriginal or Torres Strait Islander people.

However, presumably 1 January was sometimes being entered when the year of birth was known but the day and month were not (in this case, the Data Guide instructions are to enter 99 for the day and month). Allowing for the Aboriginal and Torres Strait Islander component, it was estimated that this could have been the case for 197 records, but there is no way of separately identifying these.

Results of linkage

After linkage, the estimated total number of consumers was 62,341. Almost all linkage occurred within the one jurisdiction (State, Territory or Commonwealth) or between State/Territory and Commonwealth services located within the same State or Territory. However, there were 36 matches of the linkage key between States and Territories, meaning that these consumers were assumed to be using services from two different States. At least 33 of these matches appear genuine, as they were between Commonwealth or State services in geographically adjacent State or Territories, for example in the Albury/Wodonga region. The mean number of services received per consumer was 1.20 (74,929 records divided by 62,341 consumers), or 1.21 if only records with valid linkage keys are considered (71,435 records divided by 58,847 consumers).

Table A4.1 shows the number of records per linkage key. About 64% (47,736) of all records had a valid linkage key that did not match with any other record. A further 5% (3,494) of records had an invalid linkage key and thus could not be matched. This was a higher percentage than in 1999. However, the 2000 collection was the first year that Western Australia used the standard linkage key and for this State the linkage key was invalid for 22% of consumers. For the other jurisdictions, the proportion of invalid keys ranged from 0.01% for the Commonwealth to 7% for the Northern Territory.

Thus, altogether, 68% (51,230) of all records were unmatched, and as a result there were 51,230 consumers for whom there was only one record. The other 32% (23,699) of records did have at least one match and were shared between 11,111 consumers. For example, there were 19,562 records (9,781 multiplied by 2) for the 9,781 consumers who had two matching records. The number of records with the same linkage key ranged from one to six. Nearly nine-tenths of the matches found were between two records only (88% or 9,781 of 11,111).

Table A4.1: Number of records that match using the statistical linkage key and resulting number of consumers

Effect of linkage key	Records		Consumers	
	No.	%	No.	%
<i>Unmatched records</i>				
Invalid linkage key	3,494	4.7	3,494	5.6
Valid linkage key	47,736	63.7	47,736	76.6
<i>Total</i>	<i>51,230</i>	<i>68.4</i>	<i>51,230</i>	<i>82.2</i>
<i>Linked records</i>				
Two records	19,562	26.1	9,781	15.7
Three records	3,591	4.8	1,197	1.9
Four records	484	0.6	121	0.2
Five records	50	0.1	10	<0.05
Six plus records	12	<0.05	2	<0.05
<i>Total</i>	<i>23,699</i>	<i>31.6</i>	<i>11,111</i>	<i>17.8</i>
Grand total	74,929	100.0	62,341	100.0

Records with invalid linkage keys cannot, of course, be matched with any other records, so result in an over-estimate of the number of consumers. From the results of linkage among records with valid linkage keys, an estimated 636 of the records with invalid keys would be expected to show a match if they had a valid key, and the total for consumers would decrease by this number. To this can be added the estimated 197 records that are assumed to have an invalid linkage key due to the day and month of the birth date being recorded as 01/01 when they should have been missing (see above). This would mean that the total number of consumers is over-estimated by 833 or 1.4%. However, the statistical linkage key by its nature does not result in perfect matching and can result in both false matches and missed matches. Previous testing of the linkage key indicated a false match rate of under 1% (AIHW 1998b).

Methods for resolving discrepancies between linked records

When records are matched by linkage key they are assumed to then relate to the same consumer. In the majority of cases the information on matching records will be the same. However, in some cases the other information on two or more matching records is not entirely consistent; that is, for some items the values recorded will differ between records. For example, of two records with the same linkage key, one may record the consumer as Aboriginal and with primary disability intellectual, the other as being neither Aboriginal nor Torres Strait Islander and having primary disability physical. Depending upon the item in question, this may be due to differences in opinion, judgment or quality of information, or possibly because of miscoding on one record.

In order to produce any tabulations or analysis of items by consumer, it is necessary to reconcile such discrepancies by some method that is consistent for each item. The processes used for each item and the rationale for them are described in the following sections.

Missing and 'not known' values

A general principle used in all cases is that valid values for an item should take precedence over missing or 'not known' values. This is on the basis that valid values imply the relevant information was known and could be given, whereas missing/not known values imply that the information was either unknown or would not be disclosed. Thus, where there are only two records and one has a particular item missing or not known, the value on the other record will be the one assigned to the consumer.

Age and sex

Age and sex have no discrepancies because they are both part of the linkage key, so if the linkage key matches then so must these two variables. The only exception is where records with missing sex have been linked to records with non-missing sex (see under 'Methods for the linkage of records' on page 103). In these cases the non-missing sex value is used.

Country of birth

- Order of precedence: Other specified country, non-English-speaking country unspecified, other English-speaking country, Australia.
- A specific code is given precedence over a more general code (eg Vietnam rather than South-East Asia or non-English-speaking country unspecified).
- Where two or more other countries are specified, precedence is given to responses from services in order of service type order 1 (Box A4.1).
- For any remaining cases precedence given in order of frequency of occurrence in the MDS (no cases in 2000).

Rationale

Australia may be used as the default response. Where two or more other countries are recorded (rare), then services with greater contact with the consumer can more reasonably be expected to have more accurate and complete information about the consumer.

Indigenous status

- Order of precedence: Aboriginal and Torres Strait Islander, Torres Strait Islander, Aboriginal, not Aboriginal or Torres Strait Islander.

Rationale

Anyone who is identified at some time as Aboriginal or Torres Strait Islander is likely to be so. People of both Aboriginal and Torres Strait Islander descent are more likely to be recorded as either one, rather than vice versa. Torres Strait Islanders are more likely to be recorded as Aboriginal than vice versa.

Method of communication

- Order of precedence: Little or no effective communication, Sign language, Spoken language.

Rationale

If the person is regarded as having 'Little or no effective communication', then it is assumed that this is the case in some contexts even if not apparent in others. Hence, this is a more accurate overall assessment of the difficulties of communication for the person. The same assumption applies to a lesser degree to the use of sign language or other non-spoken communication rather than spoken language.

Main language spoken at home

- Order of precedence: Other specified, 'Other language', English.
- Where two or more non-English languages are specified, precedence is given to responses in order of service type order 1 (Box A4.1).
- For any remaining cases, precedence is given in order of frequency of occurrence in the MDS (no cases in 2000).

Rationale

English may be used as the default response. Where two or more other languages are recorded (rare), then services with greater contact with the consumer can more reasonably be expected to have more accurate and complete information about the consumer.

Main source of income – child (under 16 years)

- Precedence given to 'Yes'.

Rationale

No may be used as the default response.

Main source of income – adult (16 years and over)

- Precedence is given to responses in order of service type order 2 (Box A4.2).
- For any remaining cases, precedence given in order of frequency of occurrence in the MDS. There were 12 such cases in 2000, of which 8 resulted in selecting DSP ahead of another response.

Rationale

Employment services are expected to have more accurate and complete information on income, followed by services with greater contact with the consumer.

Living arrangements/accommodation type

- Precedence is given to responses in order of service type order 1 (Box A4.1).
- For any remaining cases, precedence is given in order of frequency of occurrence in the MDS (13 cases in 2000).

Rationale

Services with greater contact with the consumer can more reasonably be expected to have more accurate and complete information about the consumer.

Disability groups (all disabilities including primary disability)

- Count all disabilities recorded.
- If one or more records include deafblind and other records include vision and/or hearing, then only deafblind is included. For these cases, if primary disability is vision or hearing, it is also changed to deafblind. (Note that it is possible to record both vision and hearing but not deafblind as the latter specifically refers to 'dual sensory impairments causing severe restrictions in communication, and in the ability to participate in community life'.)

Rationale

Services commonly indicate only those disabilities that relate to their target group and not others. The majority of discrepancies are due to differences in the numbers of disability groups recorded, rather than one disability group being recorded instead of another.

Primary disability group

- For three or more records, where possible take the majority value.
- For two records or where there is no majority, take the disability group that is most often recorded overall, counting all primary and other significant disabilities (provided it has been recorded as primary at least once).
- For remaining cases, precedence is given to responses in order of service type order 1 (Box A4.1).
- For any remaining cases, precedence is given in order of frequency of occurrence in the MDS. (There were 27 such cases in 2000, which were mostly resolved to intellectual (18), as well as physical (3), autism (3), specific learning/ADD, speech and neurological (1 each).

Rationale

Primary disability will often depend upon which of a number of disability groups is chosen as being the most important. This is largely a matter of judgment. For unresolved cases, services with greater contact with the consumer can more reasonably be expected to have more accurate and complete information about the consumer.

Other significant disabilities

- Includes all disability groups other than that selected for primary disability.

Rationale

Follows from above.

Other significant disabilities – whether present

- If, based on the above, more than one disability group is present, set to 'Yes'.
- If only one disability group is present, then take in order of precedence 'Yes', 'No', 'Not known', 'Not stated'. If set to 'Yes', this means that the other disability groups present are unknown (as for consumers with no matching records).

Rationale

Follows from above.

Areas of need for support or assistance

- For working, precedence is given to responses in order of service type order 2 (Box A4.2).
- For all other areas, precedence is given to responses in order of service type order 1 (Box A4.1).
- For remaining cases, the higher median value is taken. (In 2000, this ranged from 48 cases for working to 106 cases for mobility.)

Rationale

Services with greater contact with the consumer can more reasonably be expected to have more accurate and complete information about the consumer, except for the area of Working for which employment services can be expected to have the most accurate information.

Box A4.1: CSDA MDS service type order 1 for resolution of discrepancies after statistical linkage

Order	MDS code	Service type
Accommodation		
1	1.01	<i>Institutional accommodation/large residential</i>
2	1.02	<i>Hostels</i>
3	1.03	<i>Group homes</i>
4	1.04	<i>Attendant care</i>
5	1.05	<i>Outreach support/other 'in-home' living support</i>
6	1.06	<i>Alternative family placement</i>
7	1.07	<i>Other accommodation support services</i>
Case management		
8	2.08	<i>Behaviour intervention/specialist intervention services</i>
9	2.07	<i>Family/individual case practice/management services</i>
10	2.10	<i>Brokerage/direct funding/individual support packages</i>
Community access and employment		
11	3.02	<i>Day programs (post-school options/social and community support)</i>
12	3.03	<i>Other community access and day programs</i>
13	3.01	<i>Continuing education/independent living training/adult training centre</i>
14	5.02	<i>Supported employment</i>
15	5.03	<i>Open and supported employment</i>
16	5.01	<i>Open employment</i>
Counselling, intervention and therapy		
17	2.09	<i>Counselling: individual/family/group</i>
18	2.13	<i>Resource teams/regional teams</i>
19	2.06	<i>Therapy services</i>
20	2.04	<i>Early childhood intervention</i>
Respite		
21	4.02	<i>Centre-based respite/respite homes</i>
22	4.01	<i>Own home respite</i>
23	4.03	<i>Host family/peer support respite</i>
24	4.04	<i>Other respite/combination/flexible</i>
Other community support		
25	2.05	<i>Recreation/holiday programs</i>
26	2.11	<i>Mutual support/self-help groups</i>
27	2.14	<i>Other community support services</i>

Box A4.2: CSDA MDS service type order 2 for resolution of discrepancies after statistical linkage

Order	MDS code	Service type
Employment		
1	5.02	Supported employment
2	5.03	Open and supported employment
3	5.01	Open employment
Accommodation		
5	1.01	Institutional accommodation/large residential
5	1.02	Hostels
6	1.03	Group homes
7	1.04	Attendant care
8	1.05	Outreach support/other 'in-home' living support
9	1.06	Alternative family placement
10	1.07	Other accommodation support services
Case management		
11	2.08	Behaviour intervention/specialist intervention services
12	2.07	Family/individual case practice/management services
13	2.10	Brokerage/direct funding/individual support packages
Community access		
14	3.02	Day programs (post-school options/social and community support)
15	3.03	Other community access and day programs
16	3.01	Continuing education/independent living training/adult training centre
Counselling, intervention and therapy		
17	2.09	Counselling: individual/family/group
18	2.13	Resource teams/regional teams
19	2.06	Therapy services
20	2.04	Early childhood intervention
Respite		
21	4.02	Centre-based respite/respite homes
22	4.01	Own home respite
23	4.03	Host family/peer support respite
24	4.04	Other respite/combination/flexible
Other community support		
25	2.05	Recreation/holiday programs
26	2.11	Mutual support/self-help groups
27	2.14	Other community support services