



Australian Government

**Australian Institute of
Health and Welfare**

People using both Disability Services and Home and Community Care in 2010–11

Technical report

DATA LINKAGE SERIES NO. 17



Australian Government

**Australian Institute of
Health and Welfare**

*Authoritative information and statistics
to promote better health and wellbeing*

DATA LINKAGE SERIES

Number 17

People using both Disability Services and Home and Community Care 2010–11: technical report

Australian Institute of Health and Welfare
Canberra

CSI 19

The Australian Institute of Health and Welfare is a major national agency which provides reliable, regular and relevant information and statistics on Australia's health and welfare. The Institute's mission is authoritative information and statistics to promote better health and wellbeing.

© Australian Institute of Health and Welfare 2014



This product, excluding the AIHW logo, Commonwealth Coat of Arms and any material owned by a third party or protected by a trademark, has been released under a Creative Commons BY 3.0 (CC-BY 3.0) licence. Excluded material owned by third parties may include, for example, design and layout, images obtained under licence from third parties and signatures. We have made all reasonable efforts to identify and label material owned by third parties.

You may distribute, remix and build upon this work. However, you must attribute the AIHW as the copyright holder of the work in compliance with our attribution policy available at <www.aihw.gov.au/copyright/>. The full terms and conditions of this licence are available at <<http://creativecommons.org/licenses/by/3.0/au/>>.

Enquiries relating to copyright should be addressed to the Head of the Media and Strategic Engagement Unit, Australian Institute of Health and Welfare, GPO Box 570, Canberra ACT 2601.

This publication is part of the Australian Institute of Health and Welfare's Data linkage series. A complete list of the Institute's publications is available from the Institute's website <www.aihw.gov.au>.

ISSN 1833-1238

ISBN 978-1-74249-562-0

Suggested citation

Australian Institute of Health and Welfare 2014. People using both Disability Services and Home and Community Care 2010–11: technical report. Data linkage series no. 17. CSI 19. Canberra: AIHW.

Australian Institute of Health and Welfare

Board Chair

Dr Andrew Refshauge

Director

David Kalisch

Any enquiries about or comments on this publication should be directed to:

Media and Strategic Engagement Unit

Australian Institute of Health and Welfare

GPO Box 570

Canberra ACT 2601

Tel: (02) 6244 1032

Email: info@aihw.gov.au

Please note that there is the potential for minor revisions of data in this report. Please check the online version at <www.aihw.gov.au> for any amendments.

Contents

- Acknowledgments..... iv
- Abbreviations..... v
- Summary vi
- 1 Introduction.....1
- 2 Linkage process2
- 3 The data.....3
 - 3.1 Home and Community Care3
 - 3.2 Disability Services7
- 4 Linkage results.....10
- References17
- List of tables18

Acknowledgments

This report was prepared by Ms Rosemary Karmel and Dr Phil Anderson of the Australian Institute of Health and Welfare. Assistance was provided by Ms Felicity Murdoch and Ms Genevieve Whitlam.

Disability Services data for linkage was prepared by Mr Brendan Brady, and Ms Evon Bowler and Mr Mark Cooper-Stanbury provided advice on the Home and Community Care data.

The Standing Council on Community and Disability Services Advisory Council funded the report, and representatives from the Disability Policy and Research Working Group and the Research, Evaluation and Data Working Group provided valuable advice and comment.

Abbreviations

AIHW	Australian Institute of Health and Welfare
DS	Disability Services program
FMR	estimated false match rate
HACC	Home and Community Care program
KBL	key-based linkage
m_tf	estimated marginal trade-off
MDS	minimum data set
NMDS	national minimum data set
SLK	statistical linkage key

Symbols

–	nil or rounded to zero
..	not applicable
n.a.	not available
n.p.	not publishable because of small numbers, confidentiality or other concerns about the quality of the data

Summary

The purpose of this project is to show the feasibility of linking two community service program data sets, and to provide information about the characteristics and service use of the users of both programs (at the national level) for policy makers involved in developing and monitoring needs-based services. The two government programs chosen for this purpose were Disability Services (DS) and Home and Community Care (HACC), as both have a national minimum data set describing the client population and services provided. It is also expected that a group of clients would access both services.

The purpose of this paper is to describe the linkage process used to identify the extent of joint DS and HACC program use, and the methods used to derive data items for analysis. A detailed discussion of the DS and HACC programs and analysis of the linked data are presented in the companion AIHW report *People using both Disability Services and Home and Community Care 2010–11* (AIHW 2014), released with this report.

In both the DS and HACC data collections, clients are identified via the statistical linkage key SLK-581 which consists of five letters of name, full date of birth, and sex. While SLK-581 generally distinguishes well between individuals, it is possible for two individuals to have the same key. Hence, use of SLK-581 protects the privacy of individuals as they cannot be re-identified from the key. Data linkage between DS and HACC clients was undertaken using key-based linkage centred around SLK-581. This method maximises the value of the SLK-581 for linkage and has been used previously in a number of projects.

In 2010–11, there were 314,000 SLK-581 person identifiers in the DS data collection and 934,000 in that for HACC. Missing components for SLK-581 can mean that there is insufficient data for data linkage. Just over 8% of DS person identifiers and 6% of HACC person identifiers had insufficient or unreliable data for matching. As a consequence, 288,000 DS clients and 874,000 HACC clients were included in the key-based linkage process.

Overall, there were 55,000 matches between DS and HACC clients. These accounted for 17.5% of all DS clients and 19.0% of matchable DS clients (that is, those with sufficient information to be included in the linkage process). Among the much larger set of HACC clients, 5.9% of all HACC clients and 6.3% of matchable HACC clients matched to DS clients. As expected from the different target groups of the two programs, HACC clients who were also DS clients tended to be younger than other clients. On the other hand, the relatively small group of DS clients aged 65 and over were more likely than others to be accessing both services.

When records are matched by SLK-581, they are assumed to relate to the same service user. In the majority of cases, demographic information on these records will be the same. However, even within data sets, in some cases some information on two or more matching records may be inconsistent. Hence, a range of edits were carried out to ensure consistency of client demographic data within the two data sets before matching and analysis.

Just as data collected on client characteristics at different times for the same data collection can vary, so, too, can information collected on client characteristics for different data collections. Hence, differences in distributions derived for people accessing both DS and HACC services – using information from the DS NMDS and HACC MDS, respectively – may point to problems in collecting and recording some data items consistently. Indigenous status appears to be one such variable in the current analysis.

1 Introduction

The purpose of this project is to demonstrate the feasibility of linking two community service program data sets, and to provide information about the characteristics and service use of the users of both service programs. This information will be of use to policy makers involved in developing and monitoring needs-based services that cross the two programs.

There are two main project objectives:

- to further the knowledge, experience and capability to conduct data linkage across community service administrative data collections
- to analyse the pathways and characteristics of clients who are common to more than one service sector in the same time period.

The two government programs chosen for this project were Disability Services (DS) and the Home and Community Care program (HACC). They were chosen as they both have a national minimum data set (NMDS) describing the services provided and client population and it was expected there would be a group of clients who would access both services. The Australian Institute of Health and Welfare (AIHW) is responsible for coordinating and supporting the collection of data for the DS NMDS. For the period in this study, the HACC MDS data for all jurisdictions except New South Wales were obtained from the National Data Repository held by the then Australian Government Department of Health and Ageing. Data for New South Wales were provided separately by the New South Wales data repository and collated with data for the other states by the AIHW.

The purpose of this paper is to describe the linkage process used to identify the extent of joint DS and HACC program use. In addition, the methods used to derive a client's demographic characteristics when there is inconsistent reporting across service use records are presented. A detailed discussion of the background to the project and the DS and HACC programs, together with analysis of the linked data, is presented in *People using both Disability Services and Home and Community Care 2010–11* (AIHW 2014).

Before data linkage was undertaken for this study, approvals were obtained from required ethics committees, and permission to use the DS and HACC data was obtained from data custodians.

2 Linkage process

Data linkage between DS and HACC clients was undertaken using key-based linkage (KBL) centred around the statistical linkage key SLK-581. This key consists of five letters of name, full date of birth, and sex. This method maximises the value of the SLK-581 for linkage and has been used in a number of projects (for example, Karmel et al. 2010). The KBL process involves matching via multiple deterministic match passes, using linkage keys derived from data items available for linkage in order of key quality. Three measures – calculated for each match key – are used in this process to identify suitable linkage keys and their order of use:

- the *estimated false match rate* (FMR) for links established using the match key
- the *estimated marginal trade-off* (m_tf) between additional true and additional false matches for links established using the match key, when compared with matches made by a slightly more precise key
- a measure of *discriminating power* (expressed as %). This is the product of the unique key rates for the two data sets being linked, where the unique key rate is the proportion of records within a data set that have a unique value for the key in question .

The first two of these measures are used to identify keys to be used in the linkage process by setting upper limit cut-offs; the third determines their order of use (highest to lowest). The derivation of these measures and a more detailed description of KBL are given in Karmel et al. 2010 and AIHW 2011. Note that the number of keys selected for a linkage process depends on a range of factors, including the size of the groups being matched, the match rate and the number of variables available for inclusion in the linkage keys.

KBL matching allows for variation in reported values of match key elements for clients within a data set. For individual keys, the number of versions allowed when using a particular key is limited by max_FMR/FMR. For example, if max_FMR is set to 0.5%, up to 5 different versions (altogether) of the information for the event being matched could be considered when matching using a key with FMR = 0.1% ($0.5/0.1=5$). Versions of key elements are given a priority ranking to determine their order of use.

To link DS and HACC data, the KBL process used linkage keys based on components of the statistical linkage key SLK-581 and postcode of usual residence. The four main constituents of SLK-581, and the related KBL key components, are:

- 2nd, 3rd and 5th letters of surname (giving four components: S23, S25, S35, S235)
- 2nd and 3rd letters of given name (giving one component: F23)
- day, month and year of birth (giving three components: d, m, y)
- sex (providing one component: s)

Four key components indicating region were derived from postcode:

- pc4, pc3 pc2, pc1, where pc4 is all 4 digits of postcode, pc3 is the first 3 digits of postcode, and so on.

The KBL process for the current project used an FMR limit of 2% and an m_tf limit of 5. When linking aged care data sets, comparisons of matches obtained using KBL with name-based matches have shown that this KBL strategy produces a reasonable trade-off between false and missed matches, and results in a positive predictive value of 94%–98% (that is, percentage of matches that are correct) and sensitivity of around 93% (that is, percentage of true matches that are identified).

3 The data

3.1 Home and Community Care

In 2010–11, the HACC program was a joint Australian, state and territory government initiative under the *Home and Community Care Act 1985* (Cwlth). The Australian Government contributed approximately 60% of program funding nationally and maintained a broad strategic role, while the state and territory governments were responsible for the day-to-day administration of the program. The program funded services for older people, younger people with disability, and their carers, who live in the community, and whose capacity for independent living is at risk, or who are at risk of premature or inappropriate admission to long-term residential care facilities (DoHA 2009). In 2010–11, the bulk of home-based and community-based services for older Australians were provided under the auspices of HACC, with services ranging from needs assessment and home maintenance to nursing services.

Collection of the MDS started in January 2001. Data are collected on the characteristics of both clients and their carers, and on the services provided to both groups. All service providers in receipt of HACC funding are required to collect and provide data; data are collected progressively and aggregated for quarterly transmission. HACC MDS data are sent either to a state data repository and then to the National Data Repository, or directly to the National Data Repository. Data transmissions are validated and feedback is provided by the National Data Repository (DoHA 2009).

Counting clients

The HACC MDS does not contain a unique client identifier. Rather, it contains data items through which repeat assessments by individuals can be identified with high probability: namely, SLK-581. Previous analysis has shown that SLK-581 distinguishes well between individuals in aged care data sets (AIHW: Karmel 2005a, 2005b, 2006; AIHW: Karmel & Braun 2004).

Although not common, different people can have the same SLK-581 (0.6% in a population of 440,000) (AIHW: Ryan et al. 1999:78). The likelihood of different people having the same SLK-581 increases with the size of the population. A model fitted for the original Supported Accommodation Assistance Program statistical linkage key, which used slightly different letters of name and only year of birth, showed a quadratic relationship between size of population and number of people with duplicate keys (Karmel 2000). In 2010–11, there were 3.15 million service use records on the HACC MDS. Therefore, to reduce the likelihood of combining data for different people, as in previous studies it was decided to define a client on the HACC MDS by SLK-581, combined with the first digit of the client's postcode of usual residence. That is, essentially a client was defined by SLK-581 within a state or territory, except with New South Wales and the Australian Capital Territory being combined. (Note that, under this definition, a person who moves during the year will be identified as two clients if, on moving, the first digit of their postcode of usual residence changed.)

Records with more than one missing element of SLK-581 were considered to have insufficient information for client identification. In addition, since its inception, the date of birth '1 January' has been overly common in the HACC MDS. It has occurred at least 10 times more than would be expected from a roughly uniform distribution of births across

the year ($1/365=0.27\%$) (AIHW 2011:Table C.7). Most of these dates of birth are unlikely to have been the client's actual date of birth, and records with such dates of birth could relate to a number of different people. Consequently, all 1 January dates of birth were considered to be dummy values, and so not suitable for client identification.

Unmatchable clients

Records with insufficient information for client identification could not be used in the DS-HACC linkage process and so were 'unmatchable'. Unmatchable HACC records therefore included records with:

- more than one component of SLK-581 missing
- a 1 January date of birth
- any missing components of date of birth.

Just under 0.4% of HACC records were unmatchable due to missing SLK-581 components, and 6% were unmatchable because of a 1 January date of birth.

As stated above, unmatchable records were excluded from the HACC data set for linking to the DS data set. However, to get an indication of the number of people represented in the unmatchable records, postcode was used to provide some additional distinguishing data. Unmatchable records were therefore grouped into notional 'clients' using SLK-581 (including the missing and unreliable components) in conjunction with full postcode. (Note that some HACC clients may have had both matchable and unmatchable records; the extent of this overlap cannot be gauged.)

Using this approach, unmatchable records were grouped into a total of 60,331 unmatchable 'clients': 11,563 unmatchable records with missing SLK-581 components related to an estimated 5,161 clients, and 124,603 records with a 1 January date of birth were for an estimated 55,170 clients. The service use data for these unmatchable 'clients' were included in the analysis data to allow analysis of total program use. (Note that these numbers include a small number of clients who had no reported HACC service use in 2010–11 (see Table 4.1)).

Matchable clients

All HACC records with complete SLK-581 data were considered 'matchable' to the DS data set. For other records, the type of missing or poorly reported SLK-581 components determined whether the record could be used in the linkage process.

A small proportion (0.7%) of HACC 2010–11 quarterly records for clients had some missing elements of SLK-581. Among these, cases where records were considered to have sufficient information for client identification included those with reported postcode and complete SLK-581 data except for one of sex, letters of first name and letters of last name. (Note that if only sex were missing, this internal matching process was carried out even if postcode were also missing. This affected just 0.001% of records.)

Because each HACC agency collects its own client information, HACC clients can have missing SLK-581 components in one of their records and not in another. To reduce the effect of this on measuring client-level service use, where possible, records with sufficient data for client identification were matched deterministically back to HACC records with complete SLK-581 and postcode data.

Of the 0.7% of HACC records with some missing SLK-581 data, just under half (48%) had sufficient data for client identification, and therefore internal linkage. Depending on the

missing component, between one-fifth and one-third of quarterly records with sufficient data for internal linkage matched back to records with complete SLK-581 data. Those records that did not match were considered to relate to distinct (new) clients.

Using the above process to identify matchable HACC clients, the HACC data set for data linkage included information for 874,752 HACC care recipient clients.

Analysis data

Demographic data

A range of demographic variables were available for analysis. These included the care recipient characteristics of age, sex, country of birth, Indigenous status, living arrangement and carer availability. Carer characteristics included age, sex, relationship to care recipient and residency (co-resident or not). Because client and carer data may have been reported differently by different HACC agencies – or, in some cases, may have changed over the year – the preferred value for a characteristic for a client was derived as that most commonly reported in the HACC 2010–11 MDS. If there were no category with a clear majority, the value associated with the most recent HACC assessment was used as the preferred value. Note, however, that assessment date was missing for about 25% of HACC quarterly records. For these cases, the start date of the quarter was used instead.

In addition to variation in reporting, there were some inconsistencies in the demographic data reported in the HACC MDS. In particular, carer characteristics were not well reported. For example, while 27% of all records reported the client as having a carer, carer sex was reported in only 19% of all records. To improve data quality, a number of edits were therefore applied before deriving the preferred value. These edits are described below.

Indigenous status

- If Indigenous status were missing, the following edits were performed:
 - People reported as overseas-born and with a preferred language other than either English or an Indigenous language were assumed to be non-Indigenous.
 - People reported as Australian-born and preferring an Indigenous language were assumed to be Indigenous.
- People reported as Indigenous, but also reported as born overseas and preferring a language other than either English or an Indigenous language, were assumed to be non-Indigenous.

As a result of these edits, missing Indigenous status was recoded to non-Indigenous for 28,160 (0.9%) records and to Indigenous for 31. People reported as Indigenous were recoded as non-Indigenous in 500 records. In the analysis, Indigenous status was based on whether the client was ever identified as Indigenous after applying the above edits.

Client age

It was assumed that clients cannot be older than 115. Based on reported date of birth, client ages over 115 at 30 June 2011, or negative, were assumed to be erroneous; a new age was then derived assuming the century had been misreported (588 cases).

Carer characteristics

- It was assumed that carers should be at least 7 years old and not more than 115.

- In general, using rules similar to those implemented in the DS NMDS, it was assumed that carer age and relationship data should reflect that spouses, partners and sons/daughters-in-law should be at least 16 years old and parents should be at least 15. Also, it was assumed that clients aged over 79 do not have parent carers. Consequently:
 - carer relationship was assumed to be misreported and set to missing if:
 - the client was too young for the reported carer relationship
 - the client was too old for the reported carer relationship
 - client age was assumed to be reported accurately but the carer age was assumed to be misreported and set to missing if:
 - the carer was too young for the reported relationship
 - the client and carer age difference was inconsistent with the reported relationship.
- If the care recipient was reported as living with others and having a spouse/partner carer, carer residency, if missing, was coded as co-resident.
- If carer status was not reported or was reported as 'no carer', but either carer age and sex (13,349 records) or carer residency and relationship (5,401 records) were reported, it was assumed that there was a carer.
- If, after carrying out the above edits, carer status was 'no carer', all carer variables were set to 'not applicable'.
- If after carrying out the above edits, carer status was 'has carer':
 - carers of clients reported as living alone were coded as non-resident (including spouse/partner carers)
 - spouse carers of clients reported as living with others were coded as co-resident.

Using the above consistency edits, carer status was changed for 18,750 records. In addition, carer age was changed to missing or not applicable in 23,160 records (out of 3.15 million) and sex in 19 records. Carer relationship was set to missing or not applicable in 7,849 cases. Not counting missing values recoded to 'not applicable', residency status was changed for 74,750 records. Over half of these cases related to people living alone who had a carer. For these people, in 42,173 cases residency status reported as 'unknown' was changed to 'non-resident carer'. In a further 13,062 cases the residency status for people reported living with a spouse or partner was changed from 'unknown' to 'co-resident'. The remaining large group of changes (11,610 cases) were for people with carer status changed to 'has carer' using the above edits but with no information provided on carer residency status. For these, carer residency status was changed from 'not applicable' to 'unknown'.

To get consistent carer characteristic values, after all edits had been done, preferred values for client demographics (age, sex, country of birth, and living arrangement) were identified using the majority and date rules described at the beginning of this section. Using the same rules, the preferred value for carer availability was then selected from records for a client who reported the preferred living arrangement (that is, alone or other). If the client was identified as having a carer through this process, preferred values for other carer characteristics were also derived, again using only records reporting the preferred living arrangement. After selection of preferred values, there was a small number of inconsistencies in client and carer age and carer relationship (78 with parent and 226 with son/daughter carers); carer relationship was set to missing for these cases.

Service use data

In addition to demographic data, the HACC MDS contains information on the type and amount of services provided to clients by type of service. Information about the hours of service received by a client is collected for a range of service types including domestic assistance, social support, nursing care (at home or at a centre), allied health care, personal care, centre-based day care, other food services (excluding meals prepared elsewhere), home maintenance, respite care, counselling, assessment, case management and case planning. Quantity in terms of number is collected for several service types: meals received at home or at a centre, formal linen deliveries and one-way transport trips provided. As well, the dollar value of home modifications is recorded, as is the provision (often through lending) of goods and/or equipment, including self-care aids, support and mobility aids, communication aids, aids for reading, medical care aids, and car modification.

Annual amounts were derived by service type simply by summing across all records relating to a particular HACC client. Those clients without any types of service reported as being provided during 2010–11 were dropped from the analysis data set (704 people).

3.2 Disability Services

Under the National Disability Agreement (NDA), Australian governments fund both government and non-government agencies to deliver a range of disability support services. Services are targeted at people with disability that manifests before the age of 65; however, users of any age may receive services. The program aims to enhance the quality of life of these people and maximise the opportunity for them to participate socially and economically in the community.

Collection of the DS NMDS as an annual data collection (previously under the Commonwealth State/Territory Disability Agreement) began in 2002. Where an agency delivering services under the NDA is unable to differentiate service users and/or staff according to funding source (that is, NDA or other), it is asked to provide details of all service users and staff (for each service type). It is also important to note that the DS NMDS can record information on only one informal carer per service user.

The main counts of the DS NMDS collection are service users and service type outlets. A service type is a support activity provided to a service user (for example, accommodation support, or employment). A service type outlet is a statistical counting unit that provides a particular NDA service type at, or from, a discrete location. Data are collected, usually by organisations known as agencies, for each service type outlet (agencies may deliver one or more service types provided under the NDA).

Data on service users are collected against the service types (service type outlets) over the collection period. The level of information that must be collected varies according to the particular service type. For more information on what information is collected against each service type, see Box A1 in *Disability support services: services provided under the National Disability Agreement 2010–11* (AIHW 2012).

Counting clients

Like the HACC MDS, the DS NMDS does not contain a unique client identifier. Rather, it contains data items through which repeat assessments by individuals can be identified with high probability: namely, SLK-581. DS clients are identified using SLK-581. To link records

within the DS NMDS, the SLK-581 components of each record for a service received are compared with the SLK-581 components of all other records. Records that have matching statistical linkage keys are assumed to belong to the same individual service user and are linked.

Because SLK-581 is not a unique identifier, there is a small probability that some of the linked records do not belong to the same individual and, conversely, that some records that did not link *do* belong to the same individual. In addition, one or more components of the SLK-581 may be missing. A linkage key is considered to be invalid if any, or all, parts of the key are missing (excluding sex). Within the DS NMDS, a range of edits are routinely carried out to reduce the impact of missing linkage key components on client identification, and therefore client counts. Overall, it is expected that the number of service users is slightly overestimated (see AIHW 2012 for details).

Unmatchable clients

For a proportion of the original service user records for Victoria, the letters of the first and last names required for SLK-581 were not collected. As is the standard practice for preparing DS NMDS data for annual statistical reporting, where possible, a 'pseudo' linkage key – consisting of date of birth, sex and postcode – was used to identify clients for these cases. The use of the pseudo linkage key, and other SLK-581s with missing components, to identify clients increases the likelihood that some records could have been wrongly matched, or that some records were not matched when they should have been (AIHW 2012). Pseudo linkage keys are considered to provide a usable linkage key for the purposes of estimating DS client numbers. While all records are included in the client count for standard DS NMDS reporting, records with only a pseudo linkage key, or without the data for even a pseudo linkage key, have insufficient data to match to a HACC record, and so were unmatchable for the purposes of this study.

Overall, 5% of DS person identifiers were unmatchable due to missing name SLK-581 data (name and/or date of birth) (Table 4.1). As HACC clients with a 1 January date of birth were considered unmatchable, DS clients with such dates of birth were also unmatchable. As a consequence, a further 3% of DS person identifiers were unmatchable. Note that, as in the HACC data set, 1 January dates of birth were considerably over-represented in the DS data set. Hence, they are likely to be estimated dates and therefore unreliable for use in linkage.

Matchable clients

All DS records with complete SLK-581 data were considered 'matchable' to the HACC data set. For other records, the type of missing or poorly reported SLK-581 components determined whether the record could be used in the linkage process. Records with only sex missing in SLK-581 were matchable; those with multiple components missing or with a 1 January date of birth were unmatchable (see above). Overall, 288,203 (92%) DS person identifiers were matchable and so were included in the linkage process (Table 4.1).

Analysis data

Demographic data

A range of demographic variables were available for analysis including:

- care recipient characteristics: sex, age, disability group, support needs for participation in life areas, country of birth, Indigenous status, living arrangement, labour force status, income, individualised funding arrangements and existence of a carer
- carer characteristics: relationship of carer to care recipient and age of informal carers.

When records are matched by SLK-581 within the DS data set, they are assumed to relate to the same service user. In the majority of cases, all the demographic information on matching records will be the same. However, in some cases, the other information on two or more matching records may not be entirely consistent.

In order to produce any tabulations or analysis of items by service user, such discrepancies must be reconciled. Standard agreed processes have been designed to select the data from the most reliable source. Depending upon the nature of the item, these may involve selection on the basis of one or more of (a) the frequency of each value recorded, (b) an order of preference by the actual value of the item or (c) an order of precedence by service type of the outlets that recorded the data. A further general principle used in all cases is that valid values for an item take precedence over missing or unknown values.

Logic checks of the data are also conducted to ensure an item is consistent with the age of the service user as at 30 June (that is, the end of the reporting period). Additional logic checks are conducted between data items to ensure responses are rational and consistent.

Service use data

In addition to demographic data, the DS NMDS contains service user information on the services received within five of the seven service type groups – accommodation support, community support, community access, respite, and employment. The exceptions are ‘advocacy, information and alternative forms of communication’ and ‘other support’ (AIHW 2012). Information about the hours of service received is also collected for some service types including attendant care/personal care; in-home accommodation support; case management, local coordination and development; learning and life skills development; other community access; own home respite; centre-based respite/respite homes; host family respite/peer support respite; flexible respite; and other respite.

4 Linkage results

Using the process described in Section 2, 40 linkage keys were selected to identify matches between DS and HACC clients. All the keys included surname data and a region indicator showing at least the state or territory of the client. Twenty (20) keys excluded one of F23, day of birth, month of birth, year of birth or sex and 6 keys excluded two of these elements. In addition, a number of keys used only two letters of surname. Together, these keys enable matching when there is some variation in reporting SLK-581 for a particular client between the DS and HACC collections. They also allow matches to be made if there is limited missing SLK-581 data in one or both data sets. Note that surname data was missing for 30 HACC clients and so these clients could not be matched using the selected keys. All clients had data for at least state or territory of the client (imputed from the state or territory of the service provider, if the client's postcode was missing).

Just over 8% of DS person identifiers and 6% of HACC person identifiers had insufficient or unreliable data for matching (Table 4.1). This means that 288,203 DS clients and 874,132 HACC clients were included in the KBL linkage process. Thirty-six (36) of the 40 keys identified a total of 54,908 matches, with 4 keys not identifying any matches (Table 4.2).

The first, and therefore most precise, key accounted for 79% of all matches, and 99% of matches were identified using 10 linkage keys (Table 4.2). In addition, just 6% of matches were made using keys with an estimated FMR of 0.3% or more (Table 4.3), and 97% were made using keys with an estimated m_{tf} of 20 or more (Table 4.4).

Excluding 66 matches to HACC clients who received no services during 2010–11, the 54,842 matches between DS and active HACC clients accounted for 17.5% of all DS clients and 19.0% of matchable DS clients (Table 4.5). Among the much larger set of HACC clients, 5.9% of all HACC clients and 6.3% of matchable HACC clients matched to DS clients.

As expected from the different target groups of the two programs, HACC clients who were also DS clients tended to be younger than other clients (Table 4.6). Just over 80% of unmatched HACC clients were aged at least 65, compared with 11% of matched clients. HACC clients aged 15–24 were more likely than others to be accessing both DS and HACC, with nearly 60% of HACC matchable clients in this age group also using DS. Although only 15% of HACC matchable clients aged 45–64 matched to DS clients, this age group accounted for the largest number of matched HACC clients (33%), reflecting the size of this group relative to younger age groups. Just 1% of HACC clients aged 65 and over were also DS clients.

On the other hand, the relatively small group of DS clients aged 65 and over were more likely than other DS clients to be accessing both services: nearly half of these older DS clients matched to a HACC client, compared with under 23% for younger clients (Table 4.7). Consequently, DS clients who matched to HACC clients tended to be older than those who did not match, with only 3% of unmatched clients being aged 65 and over compared with 11% of matched clients.

Just as data collected on client characteristics by different agencies for the same data collection can vary, so, too, can information collected on client characteristics for different data collections. Values reported for some data items should remain the same over time; for example, date of birth, age and Indigenous status. Even then, differences in reported values can occur through simple error. This is seen in the age and sex distributions for the matched

clients from the two data sources: for both of these variables, there are differences of less than 0.5 of a percentage point in the various categories, and the linkage strategy was designed to allow for just such reporting errors (Table 4.8). However, large variation in distributions for a common group of people for data items that should be the same in both data sources can reflect problems in collecting and recording this information consistently. Indigenous status appears to be one such variable in the current analysis, with 5.3% of the matched clients identified as Indigenous using the DS information but 4.5% using the HACC data item.

There are also variables for which differences may be valid; for example, carer status may change over time, and so differences in reporting could reflect actual changes. However, the differences in distributions seen in Table 4.8 suggest that there may be a need for improvement in data collection practices in one or both of the data sets.

Table 4.1: DS and HACC clients, by whether matchable, 2010–11

Client group	Number	Per cent
DS clients		
<i>Unmatchable 'clients'</i>	26,049	8.3
DS person identifiers with missing letters of name ^(a)	8,651	2.8
DS person identifiers with 1 January date of birth	10,718	3.4
DS person identifiers with missing date of birth	6,680	2.1
Matchable clients	288,203	91.7
Total	314,252	100.0
HACC clients^(b)		
<i>Unmatchable 'clients'</i>	60,247	6.4
HACC person identifiers with 1 January date of birth	55,093	5.9
HACC person identifiers with insufficient data for client identification	5,154	0.6
Matchable clients ^(c)	874,132	93.6
Total	934,379	100.0

(a) All but 20 of these relate to Victorian clients. This count includes records with both missing name and date of birth.

(b) HACC clients exclude 704 person identifiers with no reported service use in 2010–11, consisting of 84 unmatchable 'clients' (including 77 with 1 January date of birth) and 620 matchable clients.

(c) Includes 30 HACC clients with missing surname data. These clients could not be matched because the keys chosen for linking all included surname data (see Table 4.2).

Table 4.2: Matches of DS and HACC clients, by linkage key and KBL pass number, 2010–11

Pass number	Linkage key	Number	Per cent	Cumulative per cent
1	S235 F23 d m y s pc4	43,384	79.0	79.0
2	S235 F23 d m y _ pc4	356	0.6	79.7
3	S35 F23 d m y s pc4	35	0.1	79.7
4	S235 F23 d m y s pc3	1,782	3.2	83.0
5	S25 F23 d m y s pc4	52	0.1	83.1
6	S35 F23 d m y _ pc4	1	—	83.1
7	S23 F23 d m y s pc4	443	0.8	83.9
8	S235 F23 d m y _ pc3	21	—	83.9
9	S235 F23 d _ y s pc4	493	0.9	84.8
10	S25 F23 d m y _ pc4	2	—	84.8
11	S23 F23 d m y _ pc4	10	—	84.8
12	S35 F23 d m y s pc3	2	—	84.8
13	S235 F23 d _ y _ pc4	12	—	84.9
14	S235 F23 d m y s pc2	2,019	3.7	88.5
15	S235 F23 _ m y s pc4	1,318	2.4	90.9
16	S25 F23 d m y s pc3	5	0.0	90.9
17	S235 F23 d m _ s pc4	673	1.2	92.2
18	S235 F23 d m _ s pc4	—	—	92.2
19	S23 F23 d m y s pc3	35	0.1	92.2
20	S235 _ d m y s pc4	884	1.6	93.8
19	S235 _ d m y s pc4	—	—	93.8
22	S235 F23 d _ y s pc3	42	0.1	93.9
23	S235 F23 d m y _ pc2	33	0.1	94.0
24	S235 F23 _ m y _ pc4	24	—	94.0
25	S25 F23 d m y _ pc3	2	—	94.0
26	S235 F23 d m _ _ pc4	27	—	94.1
20	S235 F23 d m _ _ pc4	—	—	94.1
28	S25 F23 d _ y s pc4	2	—	94.1
29	S235 _ d m y _ pc4	41	0.1	94.2
21	S235 _ d m y _ pc4	—	—	94.2
31	S23 F23 d _ y s pc4	19	—	94.2
32	S35 F23 d m y s pc2	5	—	94.2
33	S235 F23 d _ y _ pc3	1	—	94.2
34	S35 F23 _ m y s pc4	6	—	94.2
35	S235 F23 _ m y s pc3	112	0.2	94.4
36	S235 F23 d m y s pc1	2,964	5.4	99.8
37	S25 F23 d m y s pc2	13	—	99.8
38	S25 F23 _ m y s pc4	11	—	99.9

(continued)

Table 4.2 (continued): Matches of DS and HACC clients, by linkage key and KBL pass number, 2010–11

Pass number	Linkage key	Number	Per cent	Cumulative per cent
39	S23 F23 d m y s pc2	40	0.1	99.9
40	S23 F23 _ m y s pc4	39	0.1	100.0
All		54,908	100.0	..

Notes

1. Table includes 66 HACC clients with no service use reported for 2010–11 who linked to a DS client.
2. Components may not sum to the total due to rounding.

Table 4.3: Matches of DS and HACC clients, by FMR, 2010–11

Estimated FMR	Frequency	Per cent
0–<0.1	51,527	93.8
0.1–<0.2	171	0.3
0.2–<0.3	31	0.1
0.3–<0.4	3,076	5.6
0.4–<0.5	64	0.1
0.5216	39	0.1
All	54,908	100.0

Note: Table includes 66 HACC clients with no service use reported for 2010–11 who linked to a DS client.

Table 4.4: Matches of DS and HACC clients, by m_tf, 2010–11

Estimated m_tf	Number	Per cent
5–<10	242	0.4
10–<15	244	0.4
15–<20	897	1.6
20–<30	3,642	6.6
30–<50	1,835	3.3
50+	48,048	87.5
All	54,908	100.0

Notes

1. Table includes 66 HACC clients with no service use reported for 2010–11 who linked to a DS client.
2. Components may not sum to the total due to rounding.

Table 4.5: Matches by client group, 2010–11

Client group	Matched		Total	Number
	No	Yes		
Matchable DS clients	81.0	19.0	100.0	288,203
All DS clients	82.5	17.5	100.0	314,252
Matchable HACC clients	93.7	6.3	100.0	874,132
All HACC clients	94.1	5.9	100.0	934,379

Note: Table excludes 704 HACC clients with no service use reported for 2010–11.

Table 4.6: HACC matchable clients by whether matched and age group, 2010–11

Age group (years)	Matched			Matched			Number
	No	Yes	Total	No	Yes	Total	
Under 15	53.8	46.2	100.0	1.2	15.3	2.1	18,219
15–24	41.5	58.5	100.0	0.7	14.3	1.5	13,396
25–44	64.0	36.0	100.0	3.2	26.6	4.6	40,478
45–64	85.3	14.7	100.0	12.8	33.1	14.1	123,225
65 and over	99.1	0.9	100.0	82.1	10.7	77.7	678,814
Total (%)	93.7	6.3	100.0	100.0	100.0	100.0	..
Total (number)	819,290	54,842	.	874,132

Note: Age is as at 30 June 2011 using date of birth as reported on the HACC data set. Table excludes 704 HACC clients with no service use reported for 2010–11, including 66 who linked to a DS client.

Table 4.7: DS matchable clients by whether matched and age group, 2010–11

Age group (years)	Matched			Matched			Number
	No	Yes	Total	No	Yes ^(a)	Total	
Under 15	85.6	14.4	100.0	21.5	15.4	20.3	58,529
15–24	85.1	14.9	100.0	19.2	14.3	18.2	52,549
25–44	82.8	17.2	100.0	30.1	26.6	29.4	84,797
45–64	77.3	22.7	100.0	26.4	33.1	27.7	79,789
65 and over	53.5	46.5	100.0	2.9	10.6	4.4	12,539
Total (%)	81.0	19.0	100.0	100.0	100.0	100.0	..
Total (number)	233,361	54,842	..	288,203

(a) Excludes 66 HACC clients with no service use reported for 2010–11 who linked to a DS client.

Note: Age is as at 30 June 2011 using date of birth as reported on the DS data set.

Table 4.8: Concordance between DS and HACC data of key characteristics for users who accessed both DS and HACC services, 2010–11

Characteristics	DS data set		HACC data set		Difference	
	Number	Per cent	Number	Per cent	Number	Percentage point
Age (years)						
0–64	49,008	89.4	48,977	89.3	31	0.1
65+	5,834	10.6	5,865	10.7	–31	–0.1
Sex						
Male	29,630	54.0	29,410	53.6	220	0.4
Female	25,205	46.0	25,356	46.2	–151	–0.2
Not stated/not collected	7	—	76	0.1	–69	–0.1
Indigenous status						
Indigenous ^(a)	2,887	5.3	2,449	4.5	438	0.8
Non-Indigenous	50,143	91.4	50,177	91.5	–34	–0.1
Not stated/not collected	1,812	3.3	2,216	4.0	–404	–0.7
Carer status						
Carer	32,416	59.1	29,701	54.2	2,715	5.0
No carer	18,898	34.5	22,623	41.3	–3,725	–6.8
Not stated/not collected	3,528	6.4	2,518	4.6	1,010	1.8
Total users who accessed both programs	54,842	100.0	54,842	100.0

(a) Ever identified as Indigenous.

Notes

1. The HACC data set includes 1,850 people with missing age information. These people were assumed to be in the 65+ group.
2. Components may not sum to the total due to rounding.

References

- AIHW (Australian Institute of Health and Welfare) 2011. Pathways in Aged Care: program use after assessment. Data linkage series no. 10. Cat. no. CSI 10. Canberra: AIHW.
- AIHW 2012. Disability support services: services provided under the National Disability Agreement 2010–11. Disability series. Cat. no. DIS 60. Canberra: AIHW.
- AIHW 2014. People using both Disability Services and Home and Community Care 2010–11. Canberra: AIHW.
- AIHW: Karmel R 2005a. Data linkage protocols using a statistical linkage key. Cat. no. CSI 1. Canberra: AIHW.
- AIHW: Karmel R 2005b. Transitions between aged care services. Cat. no. CSI 2. Canberra: AIHW.
- AIHW: Karmel R 2006. The ins and outs of residential respite care. Cat. no. AUS 80. Canberra: AIHW.
- AIHW: Karmel R & Braun P 2004. Statistical linkage across aged care programs: an exploratory example. Welfare working paper no. 44. Canberra: AIHW.
- AIHW: Ryan T, Holmes B & Gibson D 1999. A National Minimum Data Set for Home and Community Care. Aged care series. Cat. no. AGE 13. Canberra: AIHW.
- DoHA (Australian Government Department of Health and Ageing) 2009. Home and Community Care Program Minimum Data Set, 2007–08 Annual Bulletin. Canberra: DoHA. Viewed 23 December 2010.
- http://www.health.gov.au/internet/main/publishing.nsf/Content/hacc-pub_mds_sb_07-08.htm >.
- Karmel R 2000 (unpublished). Duplicates in the SAAP linkage key. Report prepared for the Data and Research Advisory Committee for the Supported Accommodation Assistance Program. Canberra: AIHW.
- Karmel R, Anderson P, Gibson D, Peut A, Duckett SJ & Wells Y 2010. Empirical aspects of record linkage across multiple data sets using statistical linkage keys: the experience of the PIAC cohort study. BMC Health Services Research 10:41.

List of tables

Table 4.1: DS and HACC clients, by whether matchable, 2010–11.....12

Table 4.2: Matches of DS and HACC clients, by linkage key and KBL pass number,
2010–1113

Table 4.3: Matches of DS and HACC clients, by FMR, 2010–11.....14

Table 4.4: Matches of DS and HACC clients, by m_tf, 2010–1114

Table 4.5: Matches by client group, 2010–11.....15

Table 4.6: HACC matchable clients by whether matched and age group, 2010–11.....15

Table 4.7: DS matchable clients by whether matched and age group, 2010–11.....15

Table 4.8: Concordance between DS and HACC data of key characteristics for
users who accessed both DS and HACC services, 2010–11.....16

This report describes the linkage process used to identify the extent of joint use of Disability Services and the Home and Community Care program. The methods used to derive analysis variables are also described. Under one-fifth of Disability Service users also used HACC in 2010–11. The analysis of joint program use is presented in the companion report *People using both Disability Services and Home and Community Care 2010–11*.