THE UNIVERSITY OF
MELBOURNE

Methodology Paper

Identification of people with disability in linked administrative data

9 December 2021

Dr Zoe Aitken, Dr Nicola Fortune, Ms Lauren Krnjacki, Dr Samia Badji, Dr George Disney, Professor Anne Kavanagh

Table of Contents

List of Abbreviations

Abbreviation	Description
ABI	Acquired brain injury
ABS	Australian Bureau of Statistics
ADHC	Ageing, Disability and Home Care
AIHW	Australian Institute of Health and Welfare
AL	Activity Limitations
BF	Body Functions
DOMINO	Data Over Multiple Individual Occurrences
DSP	Disability Support Pension
DSS	Department of Social Services
DS NMDS	Disability Services National Minimum Dataset
EF	Environmental Factors
HC	Health Conditions
ICF	International Classification of Functioning, Disability and Health
IDEA	Intellectual Disability Exploring Answers
MADIP	Multi-Agency Data Integration Project
MBS	Medicare Benefits Schedule
MCD	Medicare Consumer Directory
MeSH	Medical Subject Heading
METeOR	Metadata Online Registry
NDDA	National Disability Data Asset
NDIS	National Disability Insurance Scheme
NPV	Negative Predictive Value
NSW	New South Wales
PBS	Pharmaceutical Benefits Scheme
PH & SOMIH	Public Housing and State Owned and Managed Indigenous Housing
PPV	Positive Predictive Value
PR	Participation Restrictions
QLD	Queensland
SA	South Australia
SDAC	Survey of Disability, Ageing and Carers
SHSC	Specialist Homelessness Services Collection
VIC	Victoria

Glossary of terms

Term	Description
Sensitivity	The probability that a derived indicator correctly identifies people with disability compared to a ‘gold standard’. It is calculated as the proportion of those people identified as having disability by the ‘gold standard’ measure who were identified as having disability by the derived indicator.
Specificity	The probability that a derived indicator correctly identifies people without a disability compared to a ‘gold standard’. It is calculated as the proportion of those people identified as not having disability by the ‘gold standard’ measure who were not identified as having disability by the derived indicator.
Positive predictive value	The probability that people identified as having disability by a derived indicator truly have a disability (according to the ‘gold standard’). It is calculated as the proportion of those people identified as having disability by the derived indicator who were identified as having disability by the ‘gold standard’ measure.
False positives	People identified as having disability by the derived indicator but not by the ‘gold standard’ measure.
False negatives	People identified as having disability by the ‘gold standard’ measure but not by the derived indicator.

1. Executive summary

There is a well-recognised need for better quality data on disability in Australia. The data that are currently available largely come from surveys. The main advantage of survey data is that they are carefully designed for research and are representative of the population at large. However, there are key limitations that can be overcome by administrative data. For example, surveys are limited by sample size which means they are not sufficiently powered to analyse important subgroups (e.g., particular disability groups) and rare outcomes (e.g., health events, such as a cancer diagnosis). Experts have been calling for improved data to allow for monitoring and reporting of outcomes for people with disability. There are extensive data collected on people in an administrative context, which could potentially be used to identify and monitor outcomes for people with disability.

The National Disability Data Asset (NDDA) pilot is a government initiative which is designed to test how to best use data from administrative datasets to identify people with disability and understand their life experiences. The NDDA is currently in a pilot phase, in which five test cases are being used to demonstrate the value of an enduring data asset. This report details the methodology developed in a test case led by the Commonwealth Department of Social Services (DSS) titled “Identification of people with disability in linked administrative data for service use and outcomes reporting [housing supports]”. The first aim of this test case is to evaluate the capacity to create a comprehensive indicator of people with disability using information from linked administrative datasets.

This preliminary methodology paper addresses this aim and sets out an approach for identifying disability using linked administrative data.

The preliminary methodology paper comprises six main components:

1. A background section describing the conceptualisation and measurement of disability;
2. A literature review of international studies using administrative data to identify disability;
3. The results of the metadata analysis of available data (i.e., documentation and evaluation of the disability identification items in each administrative data source);
4. A description of the methodology used to identify disability and derive disability indicators;
5. The results of the analyses;
6. A discussion of the implications of the findings.

1.1 Conceptualisation and measurement of disability

The International Classification of Functioning, Disability and Health (ICF) is the international standard framework and classification for organising and documenting information about functioning and disability, it conceptualises functioning as an interaction between a person’s health conditions, environmental factors and personal factors.

The Survey of Disability, Ageing and Carers (SDAC) provides the ‘gold standard’ for measuring disability prevalence in Australia. The measurement approach aligns with the ICF model – the SDAC disability cohort includes people who experience activity limitations or participation restrictions associated with health conditions. While SDAC is the most detailed and comprehensive source of disability information in Australia, it is important to acknowledge that the sample is not representative of the whole population of people with disability in Australia because it does not include, for example, people living in very remote areas and discrete Aboriginal and Torres Strait Islander communities.

1.2 Recommendations from the literature review

The literature review of international studies using administrative data to identify disability yielded two key recommendations:

1. In comparison to using data from a single administrative data source, linking data from multiple sources may improve the coverage, accuracy, and representativeness of disability indicators derived from administrative data.
2. Using data from multiple time points is necessary capture the dynamics of disability.

1.3 Recommendations from the metadata analysis

The metadata analysis documented and evaluated the disability identification items in the seven administrative data sources: National Disability Insurance Scheme (NDIS), Disability Services National Minimum Dataset (DS NMDS), Data Over Multiple Individual Occurrences (DOMINO) dataset, Specialist Homelessness Services Collection (SHSC), Public Housing and State Owned and Managed Indigenous Housing (PH & SOMIH) data collection, Medicare Benefits Schedule (MBS) and Pharmaceutical Benefits Scheme (PBS).

Key recommendations from the analysis included:

1. Core data sources for disability identification should be: DS NMDS, NDIS, and DOMINO.
2. Construction of disability indicators should be limited to people aged under 65 years.
3. Data on disability group should be mapped to the six categories used for reporting SDAC data (sensory or speech; intellectual or learning; physical; psychosocial; acquired brain injury; other).
4. Data on severity should be mapped to categories used for reporting SDAC data: ‘severe or profound core activity limitation’ and ‘other’.

1.4 Methodology used to derive indicators

Based on the results of the metadata analysis and the literature review, we developed a methodology for deriving a set of disability indicators. We derived the following disability indicators:

- Disability (yes; no);
- Severe disability (severe or profound core activity limitation; other);
- Disability groups: (sensory or speech, intellectual or learning, physical, psychosocial, acquired brain injury, other), constructed to align as closely as possible with SDAC disability groups.

1.5 Key findings of the analyses

People identified as having a disability in the NDIS, DS NMDS, DOMINO, SHSC and MBS datasets were included in the ‘disability cohort’ which was linked to MADIP to assess the validity of the derived disability indicators by testing their performance against the SDAC ‘gold standard’ disability indicator and the Census disability indicator. Key findings are presented below.

1. The derived indicator underestimated the prevalence of disability compared to the SDAC disability indicator, particularly for people aged 65 years and older. The difference in prevalence is likely to be explained by (1) differences in the operational definitions of disability between SDAC and in the administrative data sources, and (2) differences in the characteristics of people in the SDAC sample and those in the administrative data population, which is based on access to support services and payments (i.e., not all people with disability receive support payments or are NDIS participants, for example).

2. Compared to people identified using the SDAC disability indicator, people identified using the derived disability indicator were more likely to be aged 25 to 64 years, men, First Nations Australians, and people with low income, and less likely to be aged 65 years and older.
3. Compared to people identified using the SDAC disability indicator, people identified using the derived disability indicator were more likely to have severe disability and psychosocial disability, and less likely to have a sensory or speech disability or ABI.
4. People who were identified as having disability in SDAC but not by the derived disability indicator (the ’false negatives’) were less likely to have an intellectual or learning, psychosocial, or ‘other’ disability, and less likely to have a severe disability compared to all people with disability identified using the SDAC disability indicator.
5. People who were not identified as having disability in SDAC but were by the derived disability indicator (the ’false positives’) were younger, more likely to be First Nations Australians and residents of New South Wales compared to all people with disability identified using the SDAC disability indicator. They were much more likely to have a psychosocial disability and less likely to have a sensory or speech, ‘other’ disability, or ABI.
6. The sensitivity of the derived disability indicator was low overall (36%) – the derived disability indicator correctly identified just over a third of people identified using the SDAC disability indicator. The low sensitivity of the derived disability indicator was expected, given the differences in the underlying operational definitions of disability of the SDAC disability indicator and the derived overall disability indicator, and the differences in the characteristics of the people in the SDAC sample compared to the administrative data.
7. There were important differences in the sensitivity of the derived overall disability indicator for different subgroups of the population. Sensitivity was higher for people aged less than 65 years, men, First Nations Australians, people with low education, and people with low income.
8. The analysis of the sensitivity of the disability group indicators provided evidence that sensitivity was very low for derived indicators of sensory or speech, ABI, and ‘other’ disabilities (<15%), and somewhat higher for derived indicators of physical (31.3%), intellectual (32.6%), and psychosocial disability (49.3%).
9. The sensitivity of the revised indicator of severe disability (46.3%) was higher than the overall disability indicator (36.0%) and was further improved by restricting the sample to people aged less than 65 years (54.9%).
10. The derived indicator had high specificity overall – it correctly identified 95.9% of people without disability according to the SDAC indicator – and for most subgroups. Specificity was lower (91-92%) for First Nations Australians, people with low education, and people with low income.
11. Despite the specificity of the derived disability indicator being high, it is important to recognise that a about a third of people identified as having a disability by the derived disability indicator were ‘false positives’ – they were not identified as having disability by the SDAC indicator. The specificity remains high because the number of ‘false positives’ (n=1233) was small relative to the number of people in SDAC who did not have a disability (n=30,061) – the denominator for the estimate of specificity.
12. The sensitivity (37.1%) and specificity (98.7%) of the derived disability indicator were slightly higher for the more inclusive version of the derived disability indicator, which included people who applied for the NDIS but had not received a plan. This suggests that the derived disability indicator should include all NDIS applicants, not just participants.
13. In the analysis of the Census data, we found that the sensitivity of the derived disability indicator (overall disability) was relatively high (59.8%), however the specificity was low (92.4%), measured against the Census disability indicator. The sensitivity of the revised indicator of severe disability was also relatively high (47.6%), though not as high as the derived overall disability indicator, and it had higher specificity (98.0%).

14. The analysis of housing outcomes, in the SDAC sample, demonstrated that the derived disability indicators (both overall and severe) produced larger estimates of inequalities between people with and without disability for all housing outcomes compared to the SDAC disability indicators. This is not surprising because the derived indicator captures a disability cohort biased towards more severe disability and socio-economic disadvantage compared to SDAC.
15. In the Census population, the derived overall disability indicator and the revised indicator of severe disability led to smaller estimates of inequalities between people with and without disability for all housing outcomes compared to the Census disability indicator, suggesting that the Census disability indicator may be capturing a cohort of people with more severe disability compared to the revised indicator of severe disability.

1.6 Recommendations from the analysis

Key recommendations from the analysis included:

1. Administrative data from DOMINO, DS NMDS (for historical analyses) and NDIS (including people who applied but were not granted plans) should be used to construct the derived disability indicators.
2. The derived disability indicators should not be used to estimate population disability prevalence. The indicators reflect services used and supports received, however not all people with disability receive payments or access services, such as the NDIS. Furthermore, the derived overall disability indicator is not based on or aligned with a conceptual definition of disability.
3. Use of the derived disability indicators for specific purposes should be guided by an understanding of the strengths and limitations of the indicators, including which subgroups of people with disability are under-represented; results should be reported with relevant caveats attached.
4. Further research is needed to:
    - refine the derived disability indicators, e.g., by using additional data sources such as education and aged care data to ensure better representation of demographic and socio-economic subgroups of the population in the NDDA;
    - quantitatively explore the extent to which the differences in the underlying populations account for the differences in the estimates of disability;
    - better understand the discrepancies between the derived disability indicator and the SDAC disability indictor – the 'false positives' and the 'false negatives' – and implications for analyses of outcomes;
    - apply analytic techniques that try to 'correct' for biases such as quantitative bias analysis;
    - assess whether the derived indicators can be used to examine rare outcomes, small population subgroups, and monitor outcomes over time;
5. Data development work is needed to improve disability identification in administrative datasets.

1.7 Conclusions

These results highlight the potential of the derived disability indicators for use in linked administrative data collections. Linked administrative data can provide insights into contacts with the disability and mainstream service system and related life outcomes (e.g., health, education and employment) for people with disability, and disability-related inequalities that cannot be obtained from administrative data.

However, it is important to understand the limitations of the derived disability indicators, which are not representative of the whole population with disability as 'disability' is understood in Australia. While this is a limitation of the data, it is to be expected. Future work should focus on understanding in more detail the population of people with disability administrative data can identify. To do this, it is vital that data on the whole Australian population, where possible, is gradually integrated into the NDDA.

Importantly, people with disability and their representative organisations must be key players in the development of disability data and statistics, and in their use.

2. Introduction

The need for better disability data has been recognised as a critical problem in Australia [1-3]. The lack of good-quality, consistent data on disability has been a major obstacle to developing evidence-based policies and interventions to improve the experiences and outcomes of people with disability and to report against Australia's obligations under the United Nations Convention on the Rights of Persons with Disability [4].

The National Disability Data Asset (NDDA) pilot was developed in response to this. Governments are testing how to best link data from administrative datasets to understand the experiences of people with disability in Australia. The NDDA is currently in an 18-month pilot phase, in which five test cases are being used to demonstrate the value of an ongoing data asset.

This methodology paper is being developed as part of the *outcomes measurement* test case led by the Commonwealth Department of Social Services (DSS). One of the aims of this test case is to evaluate the capacity to create a comprehensive indicator of people with disability using information from linked administrative datasets. This preliminary methodology paper sets out an approach for identifying disability using linked administrative data from New South Wales, Victoria, Queensland and South Australia. The methodology will be refined upon testing when the data are linked to the Multi-Agency Data Integration Project (MADIP), testing and validating the derived disability indicators against data from the 2018 Survey of Disability Ageing and Carers, the Australian 'gold standard' for measuring disability prevalence.

In this methodology paper, we present (1) a background section describing the conceptualisation and measurement of disability, (2) results of a literature review of studies that have used administrative data to identify disability, (3) results of the metadata analysis of available data, (4) a description of the methodology used to identify disability and derive disability indicators, (5) results of descriptive analyses comparing the derived indictors against published statistics for the Australian population, (6) a discussion section including the limitations of the disability indicators, and (7) conclusions and recommendations.

3. Background

This background section describes the International Classification of Functioning, Disability and Health, the definitions used to identify disability in Australian datasets, with particular emphasis on the datasets we are using in this test case, and discusses implications for deriving an indicator of disability using linked administrative data.

3.1 The International Classification of Functioning, Disability and Health

The World Health Organization's International Classification of Functioning, Disability and Health (ICF) is the international standard framework and classification for organising and documenting information about functioning and disability.

The ICF multidimensional model conceptualises functioning as a dynamic interaction between a person's health conditions, environmental factors and personal factors. The ICF model integrates the medical and social models of disability as a 'bio-psycho-social synthesis' [5].

Key ICF concepts related to functioning and disability are:

1. *Body functions* (physiological functions of body systems, including psychological functions) and *body structures* (anatomical parts of the body such as organs, limbs and their components).
2. *Impairments*: problems in body function or structure such as a significant deviation or loss.
3. *Activity*: the execution of a task or action by an individual.
4. *Participation*: involvement in a life situation.
5. *Activity limitations*: difficulties an individual may have in executing activities.
6. *Participation restrictions*: problems an individual may experience in involvement in life situations.

Functioning is an umbrella term encompassing body functions, activities and participation, while *disability* is an umbrella term for impairments, activity limitations or participation restrictions [6].

In the ICF, *Environmental factors* make up the physical, social and attitudinal environment in which people live and conduct their lives. The ICF explicitly recognises the role of environmental factors in the creation of disability, as well as the role of health conditions, and emphasises that disability cannot be inferred from a medical diagnosis alone [5].

The ICF provides a common language and conceptual basis for the description and measurement of disability. It includes separate classifications and coding structures for Body Functions, Body Structures, Activities and Participation, and Environmental Factors.

The World Report on Disability adopted the ICF as its underpinning conceptual framework and recommends its use as 'a universal framework for disability data collection' [7] (p.45). The ICF was used as the basis for collecting information on functioning in the World Health Survey, and has been used to develop question sets for collecting standardised information on disability, such as the WHO Disability Assessment Schedule. Internationally, statistical applications of the ICF include its use as a basis for capturing data in national surveys and administrative systems, for the re-analysis of existing surveys and data, and for constructing new measurement instruments [8].

In Australia, ICF-related data standards have been developed for use in administrative data collections, including a 'Standardised disability flag module', but implementation of these data standards has been limited (https://meteor.aihw.gov.au/content/index.phtml/itemId/521050). The Disability Services National Minimum Dataset has, from the early 2000s, included a 'support needs' data item based on the ICF activities and participation domains, to which many of the assessment instruments used across the country can be mapped [9]. The use of common concepts and data standards relating to disability across administrative data collections and population surveys provides a basis for relating data from different sources, for example to compare data on users of services to data on targeted population subgroups [9, 10].

3.2 Definitions of disability in key data sources in Australia

Survey of Disability, Ageing and Carers (SDAC)

The Australian Bureau of Statistics (ABS) Survey of Disability, Ageing and Carers (SDAC) is the most detailed and comprehensive source of information about disability in Australia and provides the 'gold standard' for measuring disability prevalence.

SDAC uses more than 160 questions to identify disability and distinguish different subgroups of the disability population. It includes questions relating to broad impairment types and underlying conditions causing limitations in everyday activities. The sequence of survey questions in SDAC first identifies whether an individual has conditions that have lasted or are likely to last for at least 6 months, then whether these conditions cause restrictions in their everyday life [11]. Individuals found to have restricting conditions are identified as having disability. Thus, the operational definition of disability in SDAC is 'any limitation, restriction or impairment which restricts everyday activities and has lasted, or is likely to last, for at least six months' [12].

This approach aligns with the ICF model, which conceptualises functioning and disability in terms of a dynamic interaction between a person's health conditions, environmental factors, and personal factors. The SDAC disability cohort includes people who experience activity limitations or participation restrictions associated with health conditions.

While SDAC is the most detailed and comprehensive source of disability information in Australia, it is important to acknowledge that the sample is not representative of the whole population of people with disability in Australia. The survey does not include people living in very remote areas, discrete Aboriginal and Torres Strait Islander communities, people living in hotels and short-term caravan parks, religious and educational institutions, hostels for the homeless or night shelters, and correctional institutions

Severity

Individuals identified as having disability in SDAC are asked further questions to determine the type and degree of limitation they experience. A person is identified as having a core activity limitation if they need help, have difficulty, or use aids or equipment with a core activity: mobility, self-care or communication [13]. Their overall level of core activity limitation is determined by their highest level of limitation in these activities:

- Profound limitation - always needs help with at least one core activity
- Severe limitation - needs help sometimes or has difficulty with a core activity
- Moderate limitation - no need for help but has difficulty
- Mild limitation - no need for help and no difficulty, but uses aids or has limitations

Additional questions are asked to determine whether a person has a schooling or employment restriction.

SDAC also captures information about activity limitations in other aspects of daily living: cognitive or emotional tasks, health care, reading or writing tasks, transport, household chores, property maintenance, and meal preparation [14] (p.71).

SDAC data are commonly reported for the broad group including all people with disability, and for the group of people with severe or profound core activity limitation, defined as 'sometimes or always needing help with daily self-care, mobility or communication activities' [14].

Age- and sex-specific rates of severe or profound core activity limitation from SDAC are used to estimate the 'potential population' for specialist disability services, that is, the number of people with the potential to require disability support services [15] (p.37).

Disability groups

Disability identification information from SDAC can be used to report data by 'disability group'. This is a broad categorisation based on respondents' answers to survey questions about particular impairments, activity limitations and health conditions. It is not a diagnostic grouping. Disability is categorised into the following six groups: sensory or speech; intellectual or learning; physical; psychosocial; head injury, stroke or acquired brain injury (ABI); other [14].

A person may report impairments, limitations or conditions that fall into more than one disability group, so the sum of the prevalence of all six disability groups will be greater than the total prevalence of disability. A person's main or primary disability group can be determined based on information they provide about which of their impairments, limitations or conditions causes the most problems.

Other ABS data sources

Several other ABS social surveys identify disability using a 'Short Disability Module' of 16 questions that aims to identify people with disability and their limitations and restrictions in a way that aligns with SDAC [11].

A 'Core Activity Need for Assistance' module has been included in the Australian Census since 2006, to measure the number of people 'needing help or assistance in one or more of the three core activity areas of self-care, mobility and communication, because of a long-term health condition (lasting six months or more), a disability (lasting six months or more), or old age' [16]. It is designed to align conceptually with severe or profound core activity limitation in SDAC.

Administrative data sources

In this project, we are using data from seven administrative data sources: National Disability Insurance Scheme (NDIS), Disability Services National Minimum Dataset (DS NMDS), Data Over Multiple Individual Occurrences (DOMINO) dataset, Specialist Homelessness Services Collection (SHSC), Public Housing and State Owned and Managed Indigenous Housing (PH & SOMIH) data collection, Medicare Benefits Schedule (MBS) and Pharmaceutical Benefits Scheme (PBS).

Datasets relating to disability services and supports (DS NMDS, NDIS and DOMINO) collected data on a wide range of disability characteristics. The DS NMDS was discontinued as of 30 June 2019, as state and territory-based disability support services were transitioned over to the NDIS. It used the AIHW standard for the data element 'primary disability group', defined as 'a person's impairment of body structure or function, limitation in activity, or restriction in participation chiefly responsible for the disability'. It was used to categorise disability into four broad groups: sensory/speech, intellectual/learning, physical/diverse, and psychiatric.

The NDIS records detailed information on disability to determine eligibility for inclusion in the NDIS and to monitor NDIS participants' outcomes. The NDIS provides services and supports to people with permanent and significant disability, therefore the disability eligibility information aligns with this definition of disability, and additional information is recorded on disability groups and secondary conditions.

Disability data are recorded in DOMINO to determine eligibility for benefits. The questions used to ascertain disability vary between different benefits (e.g., disability support pension, youth disability supplement, mobility allowance), therefore there is no consistent way in which disability is defined.

Some administrative datasets also contain information on disability, collected routinely in the course of managing and delivering government services and programmes that are not specific to people with disability. To improve the quality of the data describing the experiences of people with disabilities using mainstream services, the AIHW developed a 'Standardised Disability Flag', which is 'a set of questions to identify people with disabilities or long-term health conditions who experience difficulties and/or need assistance in various areas of their life'. It is designed for use in administrative data collections as a basis for providing consistent and comparable information about people with disability who access mainstream services [17].

A shortened version of this flag (including across the life areas of self-care, mobility and communication) has been included in the SHSC since July 2013. The PH & SOMIH data collection captures data on households that have a household member with disability. However, different approaches to capturing this information are used in different jurisdictions. The 'household disability status' data element in METeOR, AIHW's Metadata Online Registry, is a housing assistance data standard specified for use in PH & SOMIH and includes a suggested question: 'Do you or any other household member have a disability or health condition that limits participation in activities (such as work, cooking, gardening, self-care), or for which assistance is required, which has lasted, or is likely to last for a period of six months or more?'.

Other administrative data sources, such as MBS and PBS, capture information about individuals that may provide a basis to infer disability (e.g., diagnosed health conditions, prescription of specific drugs). However, currently only SHSC uses a standard approach for capturing information about disability in a way that enables direct comparison with SDAC data.

3.3 Implications for creating an indicator of disability using linked administrative data

No single dataset in Australia is capable of comprehensively identifying the population of people with disability. Linkage of administrative data from different sources has the potential to enable the creation of a set of disability identifiers that can be related to the SDAC definition of disability that can identify subgroups of people with disability.

However, a methodology for deriving disability indicators from linked data needs to take into account:

- the use of different definitions of disability which align to a greater or lesser extent with SDAC;
- the collection of different information on people's health conditions, impairments and functioning;
- the coverage of different datasets, including factors determining eligibility for services and supports;
- the varying frequency of collecting or reporting disability data.

These factors, which affect the quality of the data collected in each dataset, need to be taken into account when using, and interpreting the results of, a derived disability indicator.

4. Literature review

We conducted a literature review of international and Australian research relating to the use of administrative data to identify people with disability, including data related to the provision of income support (e.g., disability pension) and disability, health, education and social services.

4.1 Methods

A literature search was conducted by identifying published articles through searches of PubMed and Scopus using the initial search terms of: Disability (or impairment) AND administrative (or linked) AND identification (or data). Each term was indexed to a Medical Subject Heading (MeSH) which is a controlled vocabulary thesaurus for indexing articles, capturing other terms in that subject area (e.g., the MeSH term Disability captures "Disabled" and "Disabilities"). Websites and publications of National Statistics or National Research Data Agencies were also searched to find grey literature relating to the use of administrative data for identifying people with disability and information on national data linkage practices. We limited the search to articles published since 2010 until May 2021 and to the following five countries: Canada, United States, United Kingdom, Australia, New Zealand.

Studies were selected from the list of articles identified in the literature review in four stages. Firstly, we reviewed the titles and abstracts of all articles and identified relevant articles. Secondly, we expanded the search around each relevant article to examine all 'similar' articles (based on shared MeSH terms) and conducting backwards and forwards citation searches. Thirdly, we reviewed titles and abstracts of articles identified in the expanded search to determine whether articles were potentially relevant. Finally, we conducted full text reviews of the articles identified to determine whether they were eligible for inclusion in the literature review.

Articles were deemed eligible for inclusion in the literature review if they contained information about the identification of people with disability using administrative data. Articles were not included if the sample of people with disability were pre-identified as having a disability (for example, via a survey) and then linked to administrative data. Systematic reviews were examined to identify relevant articles but are not included in the review.

Data were extracted on: author, year, country, target disability population, datasets used for disability identification, disability identification terms, nature of validity testing, and derivation of disability indicators.

4.2 Results

Thirty-four articles were deemed eligible for inclusion from the title and abstract, of which 20 were selected from the full text review. Of these, seven from Australia [18-24], two were from Canada [25, 26], seven were from the United States [27-33], four from the United Kingdom [34-37]. There were no eligible studies from New Zealand. This search of National Statistics or National Research Data Agencies did not yield any additional information on the use of administrative or linked data for identifying disability.

Two systematic reviews were identified in the literature search [38, 39]. Brown et al 2020 reviewed articles that identified reproductive-aged women with physical and sensory disabilities in administrative health data [38]. Friedman et al (2018) reviewed the literature on using administrative data to estimate prevalence of intellectual and developmental disabilities [39]. Though these systematic reviews were not included in our literature review, we included the individual articles listed in these two articles.

The characteristics and results of studies included in the literature review are summarised below. We focus on the sources of administrative data, alignment with ICF concepts, and results of validation studies. Information extracted from each study is presented in the table in Appendix 1, grouped by country. More detail is included for the Australian studies, as these were deemed to be more relevant for informing our methodology for deriving a set of disability indicators.

Sources of administrative data

Australian studies used data from either WA [18] or New South Wales [19-24] to identify people with intellectual disability. No study used data from multiple states or national data. The studies using data from New South Wales all used data from the Disability Services Minimum Dataset (DS-MDS). Some studies used additional datasets including additional state-wide disability services data [20, 23], Ageing, Disability and Home Care (ADHC) data [19], admitted and emergency hospital data [20, 22-24], mental health ambulatory data [20, 24], education data [20, 23], public guardian data [20, 24], New South Wales Ombudsman data [20, 23], and Corrective Services [23, 24]. The study using data from WA used the Western Australian Intellectual Disability Exploring Answers database (IDEA), which contains linked de-identified data on people with intellectual disability from the Disability Services Commission client register and the Department of Education [18].

The majority of studies using data from Canada and the United States used medical or medical insurance claim-based data, using specific International Classification of Diseases (ICD) codes to identify people with disability [25-27, 29-32]. The only study that did not use medical data used pregnancy and early life longitudinal data, using ICD codes to identify disability [28]. In addition to health data sources, one study from Canada used family services and education data [25] and one study from the United States used data from a developmental disabilities surveillance programme and education data [33]. The Canadian studies only identified people with intellectual disability, while in the United States, five studies identified multiple disability types [27-30, 32], one study identified people with physical disability only [31], and one study identified children with developmental disability [33]. The studies conducted in the United Kingdom all identified people [36, 37] or children [34, 35] with intellectual or learning disability. One study used education data only [34], one used education and health data [36], one used education, health and social services data [35], and the final study identified people with intellectual disability from an intellectual disability service register [37].

Validation of derived disability indicators

We were particularly interested in articles that identified people with disability from administrative data and validated the identification via a gold standard data source or survey. Only two studies conducted analyses to validate their derived disability indicator using linked survey data [27, 32]. These studies, both conducted in the United States using medical data, demonstrated (1) that diagnosis data alone do not adequately predict functioning [32]; (2) that there was poor identification of people with mild disabilities and better identification of people with more severe disabilities [32]; and (3) that there was better identification of disability when multiple data sources were used [27, 32]. Of the Australian studies, none conducted a validation using linked data. However, five studies compared their estimates against external benchmarks and concluded that people with mild disability who are less likely to use disability services are likely to be underrepresented [18, 19, 21-24]. Previous work in Western Australia also demonstrated that the addition of education data improved accuracy for identifying intellectual disability [18].

4.3 Key learnings of relevance to the project

Accuracy of identification

The choice of data source has implications for the accuracy of identification of disability (1) because of the coverage of the data source and (2) because of the definitions used to identify disability.

- Disability-specific administrative data (e.g., disability services databases, disability-related income support data) underrepresent people with mild disability as the coverage of the data only includes those who meet defined eligibility criteria or those who choose to access services or supports, which is unlikely to be representative of all people with disability. People with mild disability, who are not in receipt of disability services or disability-related income support are unlikely to be represented in these data sources.

- Medical data also underrepresent people with mild disability or with conditions that do not require frequent medical care. This is particularly problematic for disabilities that are not the focus of active treatment, such as sensory disabilities.
- The use of education data may provide improved coverage, particularly for people with intellectual or learning disability.

The type of data source also has implications for the accuracy of identification of disability because they use different definitions of disability to collect for different purposes. There is no standard way of identifying disability, leading to difficulties in harmonising disability data from different sources. For example, a disability assessment tool to determine eligibility for benefits will use a different definition to that used to record characteristics of people who use disability services, and different to disability identified from medical diagnoses. The subgroup of people with disability identified for each of these three purposes will be very different, potentially with little overlap. Using multiple linked administrative data sources has been shown to better capture people with disability than any individual source [18, 26, 27, 38].

RECOMMENDATION: In comparison to using data from a single administrative data source, linking data from multiple sources may improve the coverage, accuracy, and representativeness of disability indicators derived from administrative data.

Temporal aspects

Data that capture a single point in time are unable to capture people's lifetime experiences of disability, which is dynamic even for people with permanent conditions, as it represents the interaction between a person's health conditions and the environment in which they live.

Data from a single point in time is problematic because:

- A disability identified in the past may not represent an individual's current disability status. Not all people with past disability will have ongoing limitations and participation restrictions, and it cannot be assumed that disability is ongoing or permanent for many disability types.
- For some types of disability, a diagnosis may occur in early life and not be reflected in current data. Some individuals with ongoing disability who do not require services or ongoing medical care are likely to be underrepresented.
- Using data from a single point in time may be appropriate for some subgroups of people with disability but not others.

While none of the studies specifically discussed the implications of the timing and frequency of data collection and reporting, it is important to investigate how the timing of different data sources will influence identification of disability to inform a best practice approach for identifying disability in administrative data.

RECOMMENDATION: Using data from multiple time points is necessary capture the dynamics of disability and may provide a more accurate and representative indicator of disability.

5. Metadata analysis

The purpose of the metadata analysis is to document and evaluate the disability identification items in the seven administrative data sources: National Disability Insurance Scheme (NDIS), Disability Services National Minimum Dataset (DS NMDS), Data Over Multiple Individual Occurrences (DOMINO) dataset, Specialist Homelessness Services Collection (SHSC), Public Housing and State Owned and Managed Indigenous Housing (PH & SOMIH) data collection, Medicare Benefits Schedule (MBS) and Pharmaceutical Benefits Scheme (PBS).

We use 'disability identification item' to mean an individual data item in an administrative dataset that can potentially be used to identify people with disability, including data items relating to disability group or severity. Disability identification items in the constituent administrative datasets will be used, alone or in combination, to derive a set of disability indicators for use in the linked dataset.

Results of the metadata analysis will be used to determine which data items are suitable for use in constructing a disability indicator, and what cautions and caveats should be attached to the use of the resulting disability indicators.

5.1 Methods

In a spreadsheet, we listed each data item of relevance for identifying disability and/or capturing information about disability. Based on information obtained from metadata (e.g., data dictionaries, data specifications in METeOR), data custodians, and other sources (e.g., service or payment eligibility criteria), the following information was extracted for each disability identification item:

- Population for whom item is available (e.g., clients of homelessness services);
- Age group of population covered (e.g., people aged over 15 years);
- Temporal basis on which data captured (e.g., annual, defined period, date of event);
- Type of disability information (identification, disability group, severity, other);
- Information captured in relation to ICF components – body function/structure (impairment) (BF); activity (limitation) (AL); participation (restriction) (PR); health condition (HC); environmental factor (EF);
- How the information is captured (e.g., self-report, service-provider assessment);
- Other relevant information (e.g., data quality or completeness).

We developed four assessment criteria to provide a structured framework to evaluate the suitability of the individual disability identification items to contribute to the disability indicators. These consisted of (1) alignment with the SDAC definition of disability; (2) coverage of population of people with disability; (3) information on disability group and severity; and (4) temporality of disability information. The four criteria were each scored 'green', 'amber' and 'red', described in Table 1.

Table 1. Description of the four assessment criteria and their scoring

Assessment criteria	Scoring
SDAC alignment To what extent does the item align with SDAC disability definitions (including cohort with severe or profound core activity limitation)?	**Green: good alignment with SDAC disability and/or 'severe or profound core activity limitation'**
	Amber: moderate alignment with SDAC disability and/or 'severe or profound core activity limitation'
	Red: poor alignment with SDAC disability definitions
Coverage What segments of the population with disability does the data set cover?	Green: Disability identification possible for broad section of population
	Amber: Identification based on eligibility for disability-related service/payment
	Red: Identification for small population subgroups (e.g., public housing tenants)
Disability group and severity Is information on disability group and severity present in a way that can be mapped to SDAC data?	Green: Disability group and severity information present and mappable to SDAC data
	Amber: Either disability group or severity information missing, and/or not readily mappable to SDAC data
	Red: No information on disability group or severity
Temporality of disability information Can information on disability identification, disability group and severity be considered current for a given time period for which data may be analysed?	Green: All disability information can be considered current for a given year (or at event date)
	Amber: Disability identification information current, but disability group and/or severity information not current for a substantial proportion of records
	Red: Disability identification information not current for a substantial proportion of records

5.2 Results

Table 2 presents key characteristics of the seven datasets and summarises the suitability of the disability identification items as assessed using the four assessment criteria. Further detail is presented in Appendix 2.

Table 2: Summary of disability identification information by dataset

Dataset	Age range	Timing reported	SDAC alignment[a]	Coverage	Disability group and severity information	Temporality of disability information
DS NMDS	<65	Start-end	Good alignment with SDAC severe or profound	Disability service users	Disability group mappable to SDAC; 'severe or profound' vs 'other'	Identification and other disability information current for each financial year
NDIS	<65 at entry into scheme	Start-end	Moderate alignment with SDAC severe or profound	NDIS participants	Disability group mappable to SDAC; NDIS severity scale roughly mappable to 'severe or profound' vs 'other'	Identification current for each NDIS plan period; other disability information current at start of most recent NDIS plan only
DOMINO	Varies by payment type	Start-end	Poor	People eligible for disability-related payments	Limited information on disability group and severity not readily mappable to SDAC	Assume disability current during payment receipt period; other disability information may not correspond to period of payment receipt
MBS	Children aged <15	Date	Poor	Patients who access relevant services from health provider	Item 289 only is specific to autism and pervasive development disorder; no severity information	Disability identification current at event date
PBS	All ages	Date	Poor	Patients who receive relevant medication prescription from health provider	Psychosocial disability; no severity information	Disability identification and disability group current at event date
SHSC	All ages, but limited data for >65 years as aged care services are available	Start-end	Good alignment with SDAC severe or profound	Homelessness support service users	No disability group; Severity ('severe or profound' vs 'moderate or mild' core activity limitation)	As at start of support period (2 or more years ago for 4% of people identified with disability)
PH & SOMIH	All ages	Start-end	Poor / varies by jurisdiction	Housing service users (no SOMIH data for NSW or VIC)	No disability group; no severity	As at start of tenancy (2 or more years ago for 84% of people identified with disability)

[a] Degree of alignment with SDAC definitions of disability and/or severe or profound core activity limitation

5.3 Strengths and limitations

We used the four criteria to identify the strengths and limitations of the disability information in each of the datasets. A summary of the strengths and limitations is given below. Further detail is presented in the table in Appendix 2.

Disability Services National Minimum Dataset (DS NMDS)

Strengths: Potential to provide good coverage of people with disability aged under 65 years who require assistance; support needs data item enables service users with severe or profound core activity limitation to be identified; data on disability group and severity present and readily mappable to SDAC data; start and end dates for receipt of services present; all disability information current for each financial year.

Limitations: Does not capture people with disability who do not need assistance; not all people with disability who need assistance access disability services; eligibility criteria vary somewhat between services and jurisdictions; transition to NDIS impacted coverage and data quality, and the DS NMDS collection was discontinued as of 30 June 2019.

National Disability Insurance Scheme (NDIS)

Strengths: Potential to provide good coverage of people with disability aged under 65 years who require assistance and/or disability-related aids or equipment; data on disability group and severity present; start and end dates for NDIS plan present; nationally consistent eligibility criteria for scheme.

Limitations: Does not capture people with disability that is not 'permanent and significant'; not all eligible people with disability apply to NDIS; severity data item categories not defined, so not possible to definitively identify NDIS participants with severe or profound core activity limitation; information on severity and disability group is most recent available for each participant, so may not be correct for earlier years of data.

Data Over Multiple INdividual Occurrences (DOMINO)

Strengths: Receipt of disability-related payments may capture some people with disability who do not access disability services (i.e., not in DS NMDS or NDIS datasets); start and end dates for receipt of benefits present; some information relevant to disability group and severity present (though not readily mappable to SDAC).

Limitations (vary by payment): for DSP, poor alignment with SDAC disability definition because eligibility defined with reference to impairments and employment restriction (not with reference to activity limitations in core and other daily activities), 2-year requirement for employment restriction (compared to 'at least 6 months' in SDAC), and eligibility further restricted by non-disability criteria (e.g., means testing); similarly, other disability-related payments have poor alignment with SDAC definitions of disability and severe or profound core activity limitation; disability group and severity information may not correspond temporally with period of payment receipt.

Medicare Benefits Schedule (MBS)

Strengths: MBS Items that specify 'autism, pervasive developmental disorder and disability' may capture children with disability who are not captured in other datasets.

Limitations: Disability only assumed (no information about activity limitation); no data on disability severity; limited information on disability group - item 289 is specific to intellectual or learning disability (autism and pervasive development disorder) whereas all others mention 'autism, pervasive developmental disorder and disability'; no start and end dates (event only).

Pharmaceutical Benefits Scheme (PBS)

Note: PBS data are not being used to identify people with disability in linked dataset, though the data are being used to distinguish between subgroups of people with disability (identifying people with psychosocial disability).

Strengths: Prescription of antipsychotics and lithium (N05A) may identify people with psychosocial disability.

Limitations: No start and end dates (event only)

Specialist Homelessness Services Collection (SHSC)

Strengths: May capture people with disability who are not captured in other datasets (i.e., not accessing disability support services or disability-related payments); basis for disability identification aligned with SDAC core activity limitation; data on severity present; start and end dates for receipt of services present.

Limitations: Only captures people who access specialist homelessness services; no data on disability group; disability identification and severity information current at start of support period, so may be out of date for long support periods.

Public Housing and State Owned and Managed Indigenous Housing (PH & SOMIH)

Strengths: May capture people with disability who are not captured in other datasets (i.e., not accessing disability support services or disability-related payments); start and end dates for receipt of services present.

Limitations: Only captures people who access housing services; disability identification varies between jurisdictions and not aligned with SDAC definitions of disability or severe or profound core activity limitation; no data on severity or disability group; disability identification information current at start of tenancy, so may be out of date for long tenancies.

5.4 Recommendations from the metadata analysis

- As most data sources are limited to identification of people with disability aged under 65 years (exceptions are SHSC and PH & SOMIH), construction of disability indicators should be limited to this age group.
- Core data sources from which to construct a disability indicator should be: DS NMDS, NDIS, and DOMINO (DSP, Youth Disability Supplement, Mobility Allowance, BSWAT Payment, Health Care Card for the child of an adult who received Carer Allowance for that child, attending a special school).
- Additional items to consider including to construct a disability indicator should be: MBS items for children with autism, pervasive developmental disorder and disability and SHSC data.
- Where available, data on disability group should be mapped to the 6 categories used for reporting SDAC data (sensory or speech; intellectual or learning; physical; psychosocial; head injury; stroke or acquired brain injury; other), plus missing/not known.
- Where available, data on severity should be mapped to three categories: severe or profound core activity limitation; other; missing/not known.
- For any indicator derived using disability information from linked administrative data sources, results of testing against population data from SDAC and the Census (including testing of linked data in MADIP) should be reported, providing detailed information about what proportion of the disability population is captured by the indicator, and the extent to which different subgroups of the disability population are captured (including by disability group, severity, and demographic characteristics).
- Data produced using the derived disability indicator should always be accompanied by a statement providing an estimate of the proportion of the disability population captured by the indicator and outlining the subgroups of the disability population not well captured by the indicator.

6. Methodology for deriving disability indicators

Based on the results of the metadata analysis and the literature review, we developed a methodology for deriving a set of disability indicators:

- Overall disability (yes; no)
- Severe disability (severe or profound core activity limitation; other)
- Disability groups (sensory or speech, intellectual or learning, physical, psychosocial, acquired brain injury, other)

Disability groups were constructed from available data items to align as closely as possible with SDAC disability groups. The mapping of the disability groupings used in NDIS, DS NMDS and DOMINO to our six disability groups is illustrated in Appendix 3. Individuals were included in a disability group in the derived indicator based on primary or secondary disabilities identified in any of the source data sets.

The methodology used to derive the disability indicators in each dataset is described in detail in Appendix 4, which contains technical specifications for each dataset including the files used to derive the disability indicators, the variables used (including coding), and details of aggregation of data. Each dataset was reworked into a common format, with one row of data per individual per financial year to reduce the size of the datasets.

Table 3 summarises the period of coverage of each data source, the number of individuals with disability identified (using the overall disability indicator), and a brief description of how people with disability were identified.

Table 3. Methodology for each dataset

	N	Period of coverage	Description
NDIS	362,698	Jul 2013 – Jun 2020	NDIS participants
DS NMDS	672,654	Jul 2010 – Jun 2019	All individuals who used disability services
DOMINO	2,922,599	Jul 2009 – Jun 2020	Individuals identified as having a medical condition, a manifest condition, incapacity to work, a terminal illness, receiving a disability-related payment, identified as a caree, or attending a special education institution
MBS	127,377	Jan 2010 – Jun 2020	Individuals with MBS codes relating to allied health services delivered to children (MBS group M10 – Autism, Pervasive Developmental Disorder and Disability Services); early intervention services; and assessment, diagnosis and development of treatment/management plans for children with disability
SHSC	78,162	Jul 2010 – Jun 2020	Clients of homelessness support services who reported having core activity limitations at the start of the support period
PH & SOMIH	113,217	Jul 2017 – Jun 2020	All people identified as having disability at the start of their housing tenancy
PBS	751,348	Jan 2010 – Jun 2020	Codes relating to prescription of antipsychotics (N05)

6.1 Creating the disability cohort

The variables from each of the datasets were merged into a single dataset, retaining records for all individuals identified as having disability using the overall disability indicator (the 'disability cohort') plus individuals identified using a number of additional disability items[1] included to enable testing the use of these additional items for identifying people with disability using MADIP linked data. Table 4 shows the number of individuals (not observations) in each dataset who contributed to the disability cohort. After merging the datasets together, the final disability cohort contained 3,429,850 individuals who were identified as having a disability using the overall disability indicator and an additional 329,170 individuals who were identified using the additional disability items.

PBS data were not used to identify individuals as having disability for inclusion in the cohort. However, PBS data were present for 2,038,741 individuals who were identified as having disability in other datasets. PBS data were retained for these individuals for testing in MADIP. 597,799 individuals had a PBS record indicating potential psychosocial disability, defined as prescription of antipsychotics.

Table 4. Number of people with disability in each dataset contributing to the disability cohort (n=3,429,850) and additional disability items (n=329,170).

	NDIS	DS NMDS	DOMINO	MBS	SHSC	PH & SOMIH	PBS
Original dataset	586,581	672,654	5,701,477	4,759,215	871,065	551,197	3,110,909
Disability cohort	362,698	672,654	2,922,599	127,377	78,162	113,217	597,799[a]
Including additional items	586,581	N/A	2,922,599	N/A	86,518	312,446	2,038,741[a]

[a] Individuals identified using PBS data were only retained in the disability cohort if they were also identified in another data source

---

[1] Additional disability items included people who applied for the NDIS but had not received a plan, people who were identified in the SHSC data as having been provided with, referred to or in need of disability services, and people in PH & SOMIH data identified as having a disability in the household file (not person file).

7. Describing the analytic cohort and comparing the derived disability indicators against population statistics

This section describes the analyses conducted and presents results, including:

- how the analytic cohort was created and the number of records contributed by each dataset;
- the overlap between different datasets including the number of individuals uniquely identified by each dataset;
- the characteristics of the analytic cohort;
- the estimated prevalence of disability and subgroups using the derived disability indicators.

7.1 Creating the cohort for analysis comparing against SDAC 2018

The analysis was conducted using data from the financial year 2018/2019 to align as closely as possible with SDAC 2018, for which data were collected between 29 July 2018 and 2 March 2019. For each dataset, records pertaining to this financial year were selected and these were then merged into a single data file to generate the analytic cohort (people identified as having a disability using the overall disability indicator). For SHSC and PH & SOMIH data, individuals were only retained in the dataset if the disability data were collected in the 2018/2019 financial year, i.e., if they had a record indicating use of homelessness services starting in the financial year (88% of data for 2018/2019 had a start date in 2018/2019, SHSC) or started a housing tenancy within the financial year (9% of data for 2018/2019, PH & SOMIH). This decision was made because, for some individuals, their service use or tenancy spanned many years (up to 66 years), therefore their disability data were not current.

Only records for people identified by the overall disability indicator were retained for the analysis. Individuals only identified using the additional disability items were not included in the cohort for analysis.

The dataset was linked to the National Death Index and individuals who had a date of death prior to 1 July 2018 were removed. The dataset was then linked to the Medicare Consumer Directory (MCD), which has the most accurate information on age and gender. This information successfully merged for 98.1% of the analytic cohort, and for those records MCD data on age and gender were used in subsequent analyses (for records not linked, age and gender from one of the source data sets were used, as described in section 7.3, below).

Although the metadata analysis suggested that the datasets are unlikely to identify people with disability aged 65 years and over, all age groups were included in the analytic cohort to quantitatively evaluate differences in disability identification by age.

The analytic cohort for the financial year 2018/2019 contained 2,498,862 individuals. Table 5 describes the number of people that contributed to the analytic cohort (after deaths removed) for each dataset.

Table 5. Number of individuals in each dataset contributing to the analytic cohort (n=2,498,862)

	NDIS	DS NMDS	DOMINO	SHSC	PH & SOMIH	MBS	PBS
Analytic cohort	264,904	190,277	2,450,921	11,906	7,773	19,793	267,581
% of cohort	10.6%	7.6%	98.1%	0.5%	0.3%	0.8%	10.7%
Severe disability	83,224	83,529	N/A	3,871	N/A	N/A	N/A
Disability group:							
Sensory or speech	43,908	23,361	169,390	N/A	N/A	N/A	N/A
Intellectual or learning	174,480	55,414	247,713	N/A	N/A	307	N/A
Physical	47,764	86,205	1,525,435	N/A	N/A	N/A	N/A
Psychosocial	46,464	97,681	1,331,417	N/A	N/A	N/A	267,581
ABI	16,083	6,560	53,825	N/A	N/A	N/A	N/A
Other	41,005	20,350	142,898	N/A	N/A	N/A	N/A

7.2 Overlap between datasets

We created a Venn diagram to illustrate the overlap between the core datasets for disability identification: NDIS, DOMINO and DS NMDS. The figure also demonstrates the overlap of the core datasets with each of MBS, SHSC and PH & SOMIH, as well as the proportion of individuals that are identified uniquely by each of these datasets.

Figure 1. Overlap between the different datasets (disability cohort n=2,498,862)

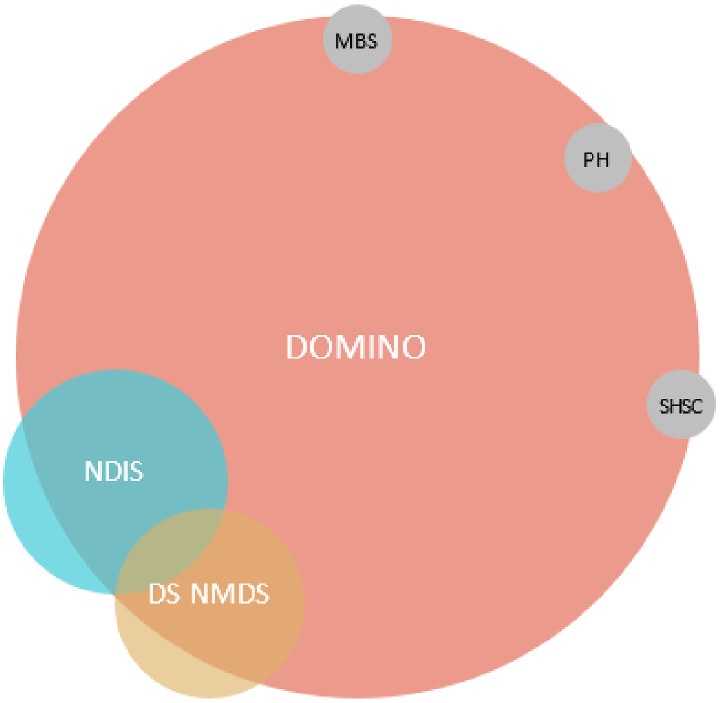

Table 6. Overlap between datasets

Core datasets	%
DOMINO only[a]	83.1%
NDIS only[a]	1.1%
DS NMDS only[a]	0.3%
DOMINO & NDIS	7.7%
DOMINO & DS NMDS	5.6%
NDIS & DS NMDS	0.1%
DOMINO & NDIS & DS NMDS	1.7%

Other datasets	Unique	Overlap
MBS	0.2%	0.8%
SHSC	0.1%	0.5%
PH & SOMIH	<0.1%	0.3%

[a] Not necessarily unique to these datasets as there may be some overlap with MBS, SHSC and/or PH & SOMIH datasets

The analysis of overlap between the datasets demonstrated that 98.1% of the cohort was identified using DOMINO (Table 5). In the core datasets, 83.1% of the cohort was identified in DOMINO only, 7.7% in DOMINO and NDIS, 5.6% in DOMINO and DS NMDS, and 1.7% in all three core datasets (Table 6). NDIS data identified 10.6% of the cohort, but only 1.1% of individuals were uniquely identified by NDIS data. DS NMDS identified 7.6% of the cohort, but only 0.3% of individuals were uniquely identified by DS NMDS.

MBS, SHSC and PH & SOMIH each identified less than 1% of the individuals with disability (Table 5), and very few were uniquely identified by these datasets (MBS: 0.2%; SHSC: 0.1%, PH and SOMIH: <0.1%, Table 6). The MBS data were intended to identify children with disability, who were less likely to be identified comprehensively in other sources, therefore the small proportion of uniquely identified individuals (0.2%) was less than expected. The small proportion of overlap (0.5%) and unique (0.1%) individuals identified using SHSC data was perhaps not surprising given that SHSC contains only a small subset of the population. The small proportion of individuals identified using PH & SOMIH data (0.3%), however, was more surprising given that the dataset contains a substantial proportion of the population. However, only 9% of individuals with housing tenancies in 2018/2019 who were identified as having a disability were retained in the dataset because their disability data was collected prior to the start of the financial year, at the start of their tenancy, so was not deemed to be current.

PBS data were not used to identify people with disability, and are therefore not included in Figure 1 or Table 6, but 10.7% of the disability cohort had a PBS record indicating potential psychosocial disability, to be tested in analyses in MADIP.

7.3 Characteristics of the analytic cohort

Table 7 presents the distribution of demographic and geographic characteristics (age, sex, Indigenous status, state of residence) and disability variables (severe disability and disability group). Where multiple data sources were available for a single variable, the value of that variable for a given record was taken from a single data set, selected according to the priority ordering recommended by AIHW (final column).

Table 7. Distribution of demographic, geographic and disability characteristics

Variable	Categories	n	%	Data source used
Age group	0-4	23,723	1.0	MCD[c]>DOMINO>NDIS>DS NMDS
	5-14	165,747	6.6	
	15-24	147,984	5.9	
	25-34	234,986	9.4	
	35-44	282,881	11.3	
	45-54	417,901	16.7	
	55-64	465,377	18.6	
	65-74	433,036	17.3	
	75+	279,298	11.2	
	Missing[a]	47,929	1.9	
Gender	Men	1,295,300	51.8	MCD>DOMINO>NDIS>DS NMDS
	Women	1,155,633	46.3	
	Missing[a]	47,929	1.9	
Indigenous status	First Nations Australians	120,904	4.8	DOMINO>NDIS>DS NMDS>SHSC>PH & SOMIH
	Not Indigenous	2,056,513	82.3	
	Missing[a]	321,445	12.9	
State of residence	NSW	835,436	33.4	DOMINO>PH & SOMIH>NDIS>DS NMDS
	VIC	648,927	26.0	
	QLD	545,589	21.8	
	SA	223,934	9.0	
	Other[b]	85,669	3.4	
	Missing[a]	159,307	6.4	
Severity	Severe	158,357	6.3	Inclusive: if any dataset reported severe
	Other	262,553	10.5	
	Missing[a]	2,077,952	83.2	
Disability group	Sensory or speech	203,905	8.2	Inclusive: multiple disability groups allowed
	Intellectual or learning	346,368	13.9	therefore % do not add up to 100%
	Physical	1,560,773	62.5	Note. Includes disability group of both
	Psychosocial	1,378,726	55.2	primary and secondary disabilities
	ABI	64,935	2.6	
	Other	193,267	7.7	
	Missing[a]	70,382	2.8	

[a] Missing for all sources

[b] Other than NSW, VIC, QLD, SA in the latest available record but identified as having resided in one of these during 2018/2019

[c] MCD, Medicare Consumer Directory

Demographic and geographic characteristics

In the analytic cohort:

- 56.1% of the cohort were aged 25 to 64 years (Table 7). There were relatively few children (7.6% aged younger than 15 years) and few older people (11.2% aged 75 years and over) despite disability being most prevalent in older ages;
- 51.8% of the cohort were men and 46.3% were women;
- 4.8% of the cohort were identified as First Nations Australians;
- 33.4% lived in New South Wales, 26.0% in Victoria, 21.8% in Queensland, and 9.0% in South Australia.

Disability characteristics

Of approximately 2.5 million people with disability identified in the cohort, only 6.3% were identified as having a severe disability (Table 7). Severity was missing for a large proportion of the cohort (83.2%) because data relating to severity in DOMINO was not mappable to SDAC categories of severity, therefore could not be used to identify people with severe disability.

The most common disability group was physical disability, which was identified for 62.5% of the cohort, followed by psychosocial disability for 55.2%. Other disability groups were less common, with 13.9% of the cohort identified as having an intellectual or learning disability, 8.2% a sensory or speech disability, 2.6% an acquired brain injury, and 7.7% categorised as ‘other’.

7.4 Prevalence of overall disability, severe disability and disability groups

To evaluate how representative the analytic cohort was of the population of people with disability as identified in SDAC, we estimated prevalence of disability and compared the number of people with disability (including disability groups and severe disability) against SDAC population statistics. For overall disability and severe disability, we also compared the prevalence of disability for different age groups, for men and women, and between the four states.

Comparing estimated prevalence using our derived disability indicators with prevalence estimated using SDAC gives insight into the extent to which subgroups of people with disability are represented in the analytic cohort.

The prevalence of disability using the derived overall disability indicator was calculated using the number of people identified in the analytic cohort as the numerator and the estimated resident population for the four States for which we had data (New South Wales, Victoria, Queensland, South Australia) at 30 June 2018 (data provided by AIHW). Population statistics of disability prevalence estimated from SDAC for the four states were retrieved using ABS Table Builder.

Prevalence of disability

The estimated overall disability prevalence was lower using the derived disability indicator (11.8%) compared to SDAC (17.8%), and the prevalence of severe disability was much lower (0.7% versus 5.8%, Table 8)

For the disability groups, for all except psychosocial disability, the estimates of prevalence were lower than SDAC estimates. There were large discrepancies for sensory or speech (1.0% versus 6.2%), acquired brain injury (0.3% versus 1.3%) and ‘other’ disabilities (0.9% versus 7.4%), and relatively smaller discrepancies for intellectual or learning disability (1.6% versus 3.0%) and physical disability (7.4% versus 11.2%). The estimated prevalence of psychosocial disability, however, was higher using the derived disability indicator compared to SDAC (6.5% versus 4.6%).

Table 8. Prevalence of disability and subgroups for the derived disability indicators and SDAC 2018

	Derived disability indicator %	SDAC disability indicator %
Overall disability	11.8	17.7
Severity		
Severe disability	0.7	5.8[a]
Disability groups		
Sensory or speech	1.0	6.2
Intellectual or learning	1.6	3.0
Physical	7.4	11.2
Psychosocial	6.5	4.6
ABI	0.3	1.3
Other	0.9	7.4

[a] 'severe or profound core activity limitation' in SDAC

Prevalence by age, gender, Indigenous status and state

The prevalence of overall disability and severe disability by age, gender, Indigenous status and state of residence is presented in Table 9.

Table 9. Prevalence of disability and severe disability by age group, gender and state of residence

		Estimated population size	Overall disability %		Severe disability %	
			Disability indicator	SDAC 2018	Disability indicator	SDAC 2018
Whole population		21,188,138	11.8	17.7	0.7	5.8
Age group	0-4	1,322,831	1.8	3.8	0.3	2.3
	5-14	2,635,093	6.3	9.5	0.8	5.6
	15-24	2,753,727	5.4	9.2	1.1	3.4
	25-34	3,163,651	7.4	7.2	0.7	1.8
	35-44	2,799,080	10.1	9.7	0.7	1.9
	45-54	2,702,661	15.5	15.6	0.9	3.6
	55-64	2,449,301	19.0	24.1	1.2	5.9
	65-74	1,153,874	22.8	38.8	0.3	9.3
	75+	1,900,310	19.1	63.4	0.0	29.0
Gender	Men	10,494,492	12.3	17.6	0.9	5.6
	Women	10,693,646	10.8	17.7	0.6	6.0
Indigenous status	First Nations Australians	586,993[a]	20.6	24.0[b,c]	1.5	8.8[b,c]
	Not Indigenous	19,877,032[a]	10.3	16.9[b,c]	0.7	4.9[b,c]
State of residence	NSW	7,980,168	10.5	16.9	0.6	5.7
	VIC	6,462,019	10.0	17.0	0.8	5.6
	QLD	5,009,424	10.9	19.1	0.7	6.2
	SA	1,736,527	12.9	19.4	0.9	6.0

[a] population estimates are from the 2016 Census rather than Estimated Resident Population 30 June 2018
[b] prevalence estimates obtained from the ABS website [40]
[c] prevalence estimates from SDAC relate to all states rather than the four states included in these data

We found that the prevalence of disability was similar for the derived disability indicators and SDAC for most age groups between 25 years and 65 years, particularly for those aged 25 to 54 years (Table 9). For example, the prevalence estimated using the derived overall disability indicator compared to SDAC was 7.4% compared to 7.2% for people aged 25 to 35 years and 19.0% compared to 24.1% for those aged 55 to 64 years. Prevalence of disability estimated using the derived indicator was substantially lower than SDAC for children and young people (0-4 years: 1.8% versus 3.8%; 5-14 years: 6.3% versus 9.5%; 15-24 years: 5.4% versus 9.2%) and very large differences were observed for people aged 65 years and above (65-74 years: 22.8% versus 38.8%; 75+ years: 19.1% versus 63.4%).

There were some differences by gender. Though the estimated prevalence of disability using SDAC was similar for men (17.6%) and women (17.7%), using the derived overall indicator of disability, the prevalence was notably higher in men compared to women (12.3% versus 10.8%).

The prevalence estimates for First Nations Australians and non-Indigenous Australians were constructed using population estimates from the 2016 Census rather than the estimated resident population for 30 June 2018, and the prevalence estimates from SDAC related to the whole Australian population rather than specific to the four states for which we had data. The estimated prevalence of disability for First Nations Australians using the derived overall disability indicator was 20.6% compared to 24.0% using SDAC.

There were only minor differences in the prevalence of disability between different states of residence. For example, using the derived overall disability indicator, the prevalence of disability was 10.0% in Victoria and 12.9% in South Australia, whereas it was estimated to be 17.0% in Victoria and 19.4% in South Australia using SDAC.

For severe disability, the derived indicator underestimated prevalence in all subgroups. This may be improved if information on disability severity available in DOMINO is included in the derived indicator of severe disability (to be tested in the linked data analysis reported in the following sections of this report).

Sensitivity analysis 1 – restricting the cohort based on age

The findings of the metadata analysis suggested that disability indicators derived using these data sources are less likely to capture people with disability aged 65 years or older. The results of our analysis were consistent with the findings of the metadata analysis. The derived overall disability indicator estimated a prevalence of disability similar to SDAC for people aged 25 to 64 years (Table 9) but underestimated the prevalence of disability in children and young people and people aged 65 years and above. Therefore, we present the results of a sensitivity analysis restricting the analytic cohort to people aged 25-64 years.

For people aged 25 to 64 years, the prevalence of disability was similar using the derived overall disability indicator (12.6%) and SDAC (13.6%, Table 10) but the prevalence of severe disability was still greatly underestimated (0.9% versus 3.2%). For the different disability groups, the prevalence of physical disability was similar to SDAC (7.9% versus 9.2%) and there was an improvement of the estimate of the prevalence intellectual or learning disability compared to the analysis using all age groups (1.2% versus 1.6% using SDAC). The prevalence of sensory or speech disability, acquired brain injury, and ‘other’ disability were substantially underestimated using the derived indicators and, in contrast, the estimated prevalence of psychosocial disability using the derived disability group indicator was almost double the prevalence estimated using SDAC (7.7% versus 4.0%).

Table 10. Prevalence of disability and subgroups for the derived disability indicators and SDAC 2018 for people aged 25 to 64 years

	Derived disability indicator %	SDAC disability indicator %
Overall disability	12.6	13.6
Severity		
Severe	0.9	3.2
Disability groups		
Sensory or speech	0.8	2.8
Intellectual or learning	1.2	1.6
Physical	7.9	9.2
Psychosocial	7.7	4.0
ABI	0.4	1.2
Other	1.1	6.5

Sensitivity analysis 2 – excluding PH & SOMIH and PBS data

One of the recommendations of the metadata analysis was to generate the disability cohort using data from five of the seven datasets only. Therefore, we present the results of a second sensitivity analysis, in which we restricted the analytic cohort to people aged 25 to 64 years and additionally constructed the analytic cohort using the five datasets recommended in the metadata analysis: NDIS, DS NMDS, DOMINO, MBS and SHSC. The results are presented for the derived overall disability indicator and a single disability group: psychosocial disability (as the other disability groups were not identified by either PH & SOMIH or PBS, therefore the results are unchanged from Table 10).

In this second sensitivity analysis, for people aged 25 to 64 years, the prevalence of disability using the derived overall disability indicator was mostly unchanged and estimated to be similar to SDAC (12.6% versus 13.6%, Table 11). Excluding PBS data had little impact on the prevalence of psychosocial disability, which was still greatly overestimated compared to SDAC (7.6% versus 4.0%).

Table 11. Prevalence of disability and subgroups for the derived disability indicators (excluding PH & SOMIH and PBS data) and SDAC 2018 for people aged 25 to 64 years

	Derived disability indicator %	SDAC disability indicator %
Overall disability	12.6	13.6
Disability groups		
Psychosocial	7.6	4.0

7.5 Discussion

In this section, we highlight the main findings, discuss the strengths and limitations of the analysis, consider the implications for using the derived disability indicators, and suggest further testing to validate the indicators.

Key findings

The analyses comparing prevalence estimated using the derived disability indicators and SDAC tells us which subgroups of people with disability may be more or less accurately represented in the data, however analysis of the data linked to SDAC is required to confirm which groups are accurately represented by the disability indicator. The preliminary results do however provide some insight into whether the derived disability indicators have the capacity to identify some subgroups of people with disability better than others.

1. We found that the prevalence of disability for people aged 25 to 64 years was similar using the derived overall disability indicator and SDAC, suggesting that, for this age group, a derived overall disability indicator from linked administrative data may provide a representative indicator of disability. The estimated prevalence was very similar for ages 25 to 54 years and slightly underestimated the prevalence of disability for those aged 55 to 64 years. However, for children (aged 0-14 years), young people (aged 15-24 years), and older people (aged 65 years and above), the derived overall disability indicator substantially underestimated the prevalence of disability and is unlikely to be suitable as a disability indicator.
2. There was also some suggestion that the derived overall disability indicator may represent men with disability better than women with disability.
3. Restricting the analytic cohort to people with disability aged 25 to 64 years for whom there was better identification of disability, prevalence of psychosocial disability was substantially overestimated using the derived disability group indicator, with an estimated prevalence more than double compared to SDAC. For all other disability groups, the derived disability group indicators underestimated prevalence.
4. For severe disability, there were large differences between the derived indicator and SDAC, with the derived indicator substantially underestimating the prevalence of severe disability compared to SDAC across all demographic subgroups. This is likely to be largely due to missing severity data in DOMINO.

The results of the analyses also provided important insights into some the limitations of the source datasets for producing derived disability indicators.

5. The severity items in DOMINO are not readily mappable to SDAC categories of severity. Therefore, we could not include information from DOMINO in our derived indicator of severe disability. With DOMINO accounting for such a large proportion of the dataset (98.1%), it is not surprising that the derived indicator substantially underestimated severe disability.
6. Data on disability group only aligned well with SDAC categories for two of the datasets: DS NMDS and NDIS. This may affect the prevalence estimates for disability groups from the derived indicator, and in part explain differences with the estimates from SDAC.
7. The results suggest that the datasets that we used to produce the derived disability indicators are not suitable for identifying the population of people with disability aged 65 years and older. The underestimate of the prevalence of disability in older age groups is not surprising, given that the disability-specific datasets mostly include people aged younger than 65 years (and the findings from the metadata analysis also recommended that the indicator should be limited to people aged younger than 65 years).
8. The results suggest that the datasets do not adequately identify children and young people with disability (younger than 25 years). It would be valuable to explore the possibility of including education data to better capture children and young people with disability.
9. Data from PBS and PH & SOMIH were assessed in the metadata analysis as less useful data sources for the purpose of deriving disability indicators. We conducted a sensitivity analysis in which we excluded these data sources, which had little effect on the results. Based on this finding, we would recommend not using these data sources for producing derived disability indicators.

Limitations of the analysis

In the NDIS data that we had access to for this project, a single record was provided for each NDIS participant, representing their most recently recorded disability data. As a result, we did not have information on prior disability characteristics, so for some individuals the information may not be correct for the2018-2019 financial year. Separate NDIS records for each financial year would provide more accurate disability information.

There were missing data on demographics (age, gender and Indigenous status) for all individuals identified as being carees of people with concession cards in our dataset (1.7% of the DOMINO data in the 2018/2019 disability cohort). Though demographic information was available on the carers, we were unable to extract demographic information about the carees. Surprisingly none of these individuals were successfully matched to the demographic data in the Medicare Consumer Directory. This may have contributed in part to the underestimate of the prevalence of disability in younger age groups.

The analytic cohort relied on linkage of records between the Medicare Consumer Directory, the National Death Index, DOMINO, NDIS, DS NMDS, MBS, SHSC and PH & SOMIH. Though the linkage rates were exceptionally high (96% for DOMINO, 99% for NDIS and 92% for DS NMDS [41]), some people will inevitably have been missed or incorrectly linked, which may lead to underrepresentation of particular subgroups of the population of people with disability.

Implications of the findings

The analysis comparing the derived disability indicators against population statistics presents promising results suggesting that the derived overall disability indicator may provide a method suitable for identifying a representative cohort of people with disability in Australia for people aged 25 to 64 years. The linked data analyses reported in the following two sections provide further insight into the validity of the derived disability indicators.

8. Linked data analysis using SDAC

Linkage of the derived disability indicators to the SDAC sample in MADIP was used to test the validity of the derived disability indicators. Analyses were conducted to assess the extent to which the derived disability indicators are representative of the population of people with disability in Australia, as identified in SDAC (the 'SDAC disability indicator'). The results of these analyses provide insight into whether the derived indicators are fit for purpose for identifying people with disability (and/or subgroups of people with disability) in linked administrative datasets for specific purposes.

This section contains a description of the methods used and the results of the analyses, including:

- creation of the AIHW disability cohort to be linked with MADIP;
- description of the linkage between datasets and the resulting sample for analysis;
- the estimated prevalence of disability using the derived disability indicators and the SDAC indicators;
- analysis of the sensitivity and specificity of the derived disability indicators against SDAC;
- description of the characteristics of SDAC participants identified as having disability by (1) both SDAC and the derived disability indicators, (2) SDAC only, and (3) the derived disability indicators only;
- sensitivity analyses conducted to inform approaches for enhancing the sensitivity and specificity of the derived disability indicators;
- analysis of indicators of disability groups;
- derivation and analysis of an indicator of severe disability.

8.1 Creating the AIHW disability cohort for linkage with MADIP

All individuals identified as having disability using the overall disability indicator and the additional disability items[2] from five datasets (DOMINO, NDIS, DS NMDS, SHSC and MBS) were included in the AIHW disability cohort that was linked with MADIP. Individuals identified as having a disability in PH & SOMIH data only were not included in the AIHW disability cohort as approvals for data linkage were not granted by data custodians. Therefore, the AIHW disability cohort linked to MADIP was a subset of the disability cohort used in the analyses in Section 7.

The AIHW disability cohort that was linked to MADIP contained 3,088,585 individuals who were identified as having a disability using the overall disability indicator and an additional 52,879 individuals who were identified using the additional disability items.

8.2 Linkage between datasets and sample for analysis

The target SDAC sample for the analysis consisted of all SDAC participants, both with and without disability, who had not died prior to 2018 (according to date of death in the Medicare Consumer Directory dataset) and who were residing in the four states for which we had data from AIHW as part of the NDDA test case: New South Wales, Victoria, Queensland and South Australia (n=40,806). The target SDAC sample was linked to the AIHW disability cohort (individuals identified by the derived overall disability indicator) to further test the performance of the derived disability indicators. This was achieved by linking both the target SDAC sample and the AIHW disability cohort to the MADIP spine.

Linkage of the SDAC sample to the MADIP population spine

Of the 40,806 individuals in the target SDAC sample, 90.9% linked to the MADIP population spine – the 'SDAC analytic cohort' – and 9.1% could not be linked to the MADIP population spine. Those who did not link were more likely to be younger, male, have higher education, low income, and not have a disability (Figure 2).

---

[2] Additional disability items included people who applied for the NDIS but had not received a plan and people who were identified in the SHSC data as having been provided with, referred to or in need of disability services

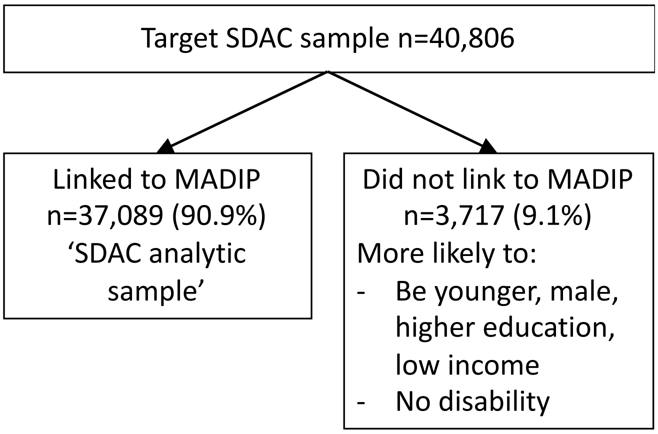

Figure 2. Flow chart illustrating linkage between SDAC target sample and MADIP population spine

The unlinked 9.1% of SDAC records has implications for the findings of the analyses. For example, because people without a disability in SDAC were less likely to successfully link with MADIP, the prevalence of disability in the linked SDAC sample is slightly overestimated compared to the target sample (17.7% versus 17.0%), and there are some differences in the age, gender, education and income distributions. However, as the difference in estimated prevalence of disability is small, the missing records are unlikely to have a substantial impact for analyses of the validity of the derived disability indicator against SDAC.

Linkage of the AIHW disability cohort to the MADIP population spine

The AIHW disability cohort was linked to the MADIP population spine. 95.0% of records were successfully linked and 5.0% could not be linked. We analysed the unlinked records to understand the reasons why they could not be linked, which included:

- 4.7% of records from the disability cohort were not included in the linkage map file supplied by AIHW to the ABS containing the AIHW identifier and the corresponding MADIP spine identifier. This group consisted of records identified only in one or more of:
  - DOMINO through information on carees of people with carers' health care concession cards (4.3%)
  - SHSC (0.4%)
  - DS NMDS (<0.1%)
- 0.2% of records that had multiple AIHW identifiers associated with a single MADIP spine identifier, or alternatively records with multiple MADIP spine identifiers associated with a single AIHW identifier. We made the decision to drop these records from the sample.
- <0.1% of records that could not be linked to the MADIP population spine.

The 5.0% of unlinked records were compared to those that did link in terms of demographics. The unlinked records were more likely to be women and First Nations Australians. We could not compare age between unlinked and linked records because the year of birth variable had been modified in the linkage process (with the whole cohort incorrectly recorded as being born in 1959 or 1960).

As well as the unlinked records from the AIHW disability cohort, there are differences in the underlying population contributing to the AIHW linked administrative datasets (Medicare Consumer Directory) and the MADIP population spine (Medicare Consumer Directory, DOMINO Centrelink administrative data, and Personal Income Tax data). Records in MADIP that were identified in the Personal Income Tax data will include people who are not eligible for Medicare, such as people who are paying taxes but are not Australian permanent residents or citizens, and these people would not be captured in the AIHW population spine. Differences in the underlying populations have implications because (1) we are linking 'cases' only (the AIHW disability cohort) and using the MADIP population spine as the denominator in our analyses, and (2) these different groups would have different characteristics, for example we would expect people who are not eligible for Medicare but paying taxes to have lower rates of disability than people identified in the Medicare Consumer Directory.

We investigated the difference in the underlying populations of the MADIP and AIHW spines using a variable in the MADIP core file indicating whether the individual had data in the Medicare Consumer Directory. We found that, in the MADIP population spine, 95.0% of records were identified in the Medicare Consumer Directory and 5.0% were identified from other sources, suggesting that 95.0% of the records in MADIP were from the same underlying population as the AIHW disability cohort. In the SDAC analytic sample, the underlying populations were more similar, with 99.0% of records in the SDAC sample identified in the Medicare Consumer Directory. Therefore, the difference in the underlying populations is likely to have had only a small impact on the results of the analysis.

Linkage of the AIHW disability cohort to the SDAC sample

The SDAC analytic sample (n=37,089) was linked to the records of the individuals in the AIHW disability cohort (n=3766) who were identified as having a disability in the 2018/2019 financial year based on the derived disability indicator (Figure 3). As 5.0% of the AIHW disability cohort were not linked to MADIP (as described above), some individuals identified in the AIHW disability cohort who participated in the SDAC survey would not have linked with the SDAC analytic sample (n=~187).

The non-linked records have implications for the findings of the analyses. Disability prevalence based on the derived disability indicator within the SDAC sample will be underestimated. Correspondingly, the true sensitivity of the derived disability indicator against SDAC will be underestimated.

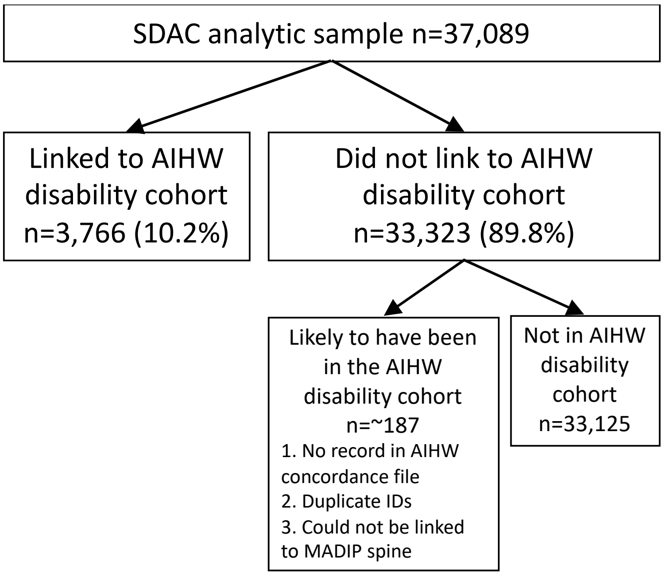

Figure 3. Flow chart illustrating linkage between SDAC analytic sample and AIHW disability cohort

8.3 Prevalence of overall disability in the SDAC sample

The prevalence of overall disability was estimated in the SDAC sample using the SDAC disability indicator and the derived disability indicator. The prevalence of disability was calculated as the number of people identified as having a disability using each of the indicators over the total sample size. Prevalence estimates were also calculated for categories of demographic, socio-economic and geographic variables, including age group, gender, Indigenous status, education, income and state of residence. Survey weights were used to account for each person's chance of selection or differences in response rates across population groups, to make the estimates representative of the Australian population [12]. The prevalence of disability in the whole sample and by demographic, socio-economic and geographic characteristics is presented in Table 12.

Table 12. Prevalence of disability using the derived disability indicator and the SDAC disability indicator

		SDAC sample	
		Derived disability indicator (%)	SDAC disability indicator (%)
Whole sample		9.8	17.7
Age group	0-4	2.7	3.9
	5-14	6.0	9.9
	15-24	5.5	10.0
	25-34	6.6	7.7
	35-44	7.8	10.0
	45-54	12.1	15.8
	55-64	15.8	24.4
	65-74	18.4	38.2
	75+	15.9	60.0
Gender	Men	10.4	18.0
	Women	9.3	17.5
Indigenous status[a]	First Nations Australians	20.5	26.7
	Not Indigenous	10.4	18.7
Education	Completed year 12	5.6	10.5
	Did not complete year 12	18.0	31.5
Income[b]	Highest 60%	5.6	13.4
	Lowest 40%	20.2	31.5
State	NSW	9.4	16.9
	VIC	9.4	17.0
	QLD	10.9	19.4
	SA	10.6	19.3

[a] There were 3245 missing observations for Indigenous status

[b] There were 12890 not applicable/not known observations for income (not asked for the whole sample)

Overall prevalence of disability was lower for the derived disability indicator compared to SDAC (9.8% versus 17.7%).

The differences between the age-specific prevalence estimates using the derived disability indicator and the SDAC indicator were larger for older age groups (e.g., the prevalence estimated using the derived disability indicator was 15.9% compared to 60.0% using the SDAC indicator for people aged 75 years and older) and smaller for younger age groups (e.g., 6.6% versus 7.7% for those aged 25 to 34 years). The prevalence of disability for First Nations Australians estimated using the derived disability indicator was 20.5% compared to 26.7% using SDAC. The difference between the derived indicator and SDAC estimates was larger for non-Indigenous people (10.4% versus 18.7%). For both education and income, the relative differences in the prevalence estimates using the derived disability indicator and the SDAC indicator were larger for people with high education and higher income compared to low education and lower income. There were no notable differences by gender or state of residence.

There were also differences between the prevalence estimates using the derived disability indicator in the SDAC sample compared to the AIHW linked data. The overall prevalence was slightly lower in the SDAC sample (10.5%) compared to the AIHW linked data (11.8%, see Table 9 in Section 7.4).

8.4 Characteristics of people with disability identified using the SDAC disability indicator and the derived disability indicator

We conducted an analysis comparing the demographic, socio-economic and geographic characteristics of individuals who were identified as having a disability using the SDAC disability indicator (n=7028) and those identified using the derived disability indicator (n=3766). The analysis was unweighted to describe the characteristics of the sample.

The results in Table 13 demonstrate that, compared to people identified by the SDAC disability indicator, the people identified by the derived disability indicator were:

- more likely to be aged between 25 and 64 years;
- less likely to be aged 65 years and older;
- more likely to be men, First Nations Australians and people with low income;
- more likely to have a severe disability, according to the revised indicator of severe disability (see Section 8.7, below);
- more likely to have a psychosocial disability;
- less likely to have a sensory or speech, ‘other’ disability, or ABI;
- had a similar proportion of people reporting intellectual and physical disability.

We also described the characteristics of people identified by the SDAC indicator but not by the derived disability indicator (n=4495) – the ‘false negatives’ – and those identified as having a disability by the derived disability indicator but not the SDAC disability indicator (n=1233) – the ‘false positives’ (see Figure 4). These groups are of interest because they represent the people who may have been incorrectly identified by the derived disability indicator. We compare the distribution of the 'false positives' and 'false negatives' against the people identified by the SDAC indicator.

Compared to people identified by the SDAC disability indicator, those people identified using only the SDAC disability indicator (and not the derived disability indicator), the ‘false negatives’, had a similar distribution of most demographic, socio-economic and geographic characteristics but there was some indication that they were more likely to have high income. They were less likely to have an intellectual or learning, psychosocial, or ‘other’ disability, and less likely to have a severe disability.

Compared to people identified by the SDAC disability indicator, people identified using the derived disability indicator only, the ‘false positives’, were more likely to be aged between 25 and 54 years and less likely to be aged 65 years and older, more likely to be First Nations Australians and residents of New South Wales. They were much more likely to have a psychosocial disability and less likely to have a sensory or speech, ‘other’ disability or ABI.

Table 13. Characteristics of people with disability identified using the derived disability indicator and the SDAC disability indicator (column %s)

		SDAC disability indicator (%) n=7028	Derived disability indicator (%) n=3766	SDAC only 'false negatives' (%) n=4495	Derived indicator only 'false positives' (%) n=1233
Age group	0-4	1.3	1.7	1.1	1.5
	5-14	6.9	7.5	6.0	5.8
	15-24	5.8	6.0	5.2	4.5
	25-34	5.1	8.1	4.8	12.9
	35-44	7.0	10.0	6.5	14.3
	45-54	10.9	15.6	9.5	19.7
	55-64	17.3	20.3	15.0	17.6
	65-74	21.2	18.9	22.1	17.4
	75+	24.4	11.9	30.0	6.3
Gender	Men	49.0	50.8	47.2	48.1
	Women	51.0	49.2	52.8	51.9
Indigenous status	First Nations Australians	3.0	4.4	2.3	4.2
	Not Indigenous	93.9	93.8	93.9	93.5
	Missing	1.9	3.1	3.8	2.3
Education	Completed year 12	37.4	35.8	39.9	41.4
	Did not complete year 12	62.6	64.2	60.1	58.6
Income	Highest 60%	30.1	23.1	37.0	33.9
	Lowest 40%	46.6	53.7	39.2	41.4
	Not applicable/not known	23.2	23.3	23.8	24.7
State	NSW	35.4	35.6	36.3	39.3
	VIC	31.7	31.1	31.8	30.2
	QLD	27.5	27.9	26.7	25.5
	SA	5.4	5.5	5.2	5.1
Disability group	Sensory or speech	35.0[a]	5.4[a]	36.8[a]	9.9[a]
	Intellectual or learning	14.4	11.9	8.7	20.1
	Physical	63.2	57.1	59.2	67.7
	Psychosocial	22.6	51.4	12.8	67.6
	ABI	6.6	0.9	4.2	3.3
	Other	40.8	4.0	32.7	7.8
Severity	Not severe	70.6	76.0	81.0	55.3
	Severe	29.4	24.0	19.0	44.7

[a] column percentages may add up to more than 100% because an individual can be included in multiple disability groups

8.5 Sensitivity and specificity of the derived disability indicator in the SDAC sample

Sensitivity describes the ability of an indicator to correctly identify people with disability. The sensitivity of the derived disability indicator was calculated as the proportion of those people identified as having disability using the SDAC disability indicator who were identified as having a disability by the derived disability indicator (Figure 4).

Specificity describes the ability of an indicator to correctly identify people without a disability. The specificity of the derived disability indicator was calculated as the proportion of those people identified as not having disability in SDAC who were not identified by the derived disability indicator.

Sensitivity and specificity are measures that help assess the accuracy of a new indicator against a 'gold standard' that is considered to provide the most accurate available measure. We used the SDAC disability indicator as the 'gold standard' measure in this analysis against which we compared the performance of the derived disability indicator.

SDAC Indicator

Derived disability indicator		Disability	No disability	
	Disability	a True positive	b False positive	Positive predictive value (PPV)= $\frac{a}{a+b}$
	No disability	c False negative	d True negative	Negative predictive value (NPV)= $\frac{d}{c+d}$
		Sensitivity = $\frac{a}{a+c}$	Specificity = $\frac{d}{b+d}$	

Figure 4. Sensitivity and specificity

We calculated the sensitivity and specificity of the derived disability indicator for the whole sample, and for categories of demographic, socio-economic and geographic variables, including age group, gender, Indigenous status, education, income and state of residence. Analyses were unweighted as they were not intended to be representative of the Australian population. The estimates of sensitivity and specificity are presented in Table 14.

We estimated that the derived disability indicator had a sensitivity of 36.0%, therefore just over a third of people identified as having a disability using SDAC were identified using the derived disability indicator (Table 14).

The sensitivity of the derived disability indicator was higher in younger age groups (ranging from 40.4% to 48.9% in age groups up to 64 years) and lower for older age groups (33.4% for people aged 65 to 74 years and 21.6% for 75 years and older). The sensitivity was higher for men (38.4%) compared to women (33.8%) and higher for First Nations Australians (52.3%) compared to non-Indigenous people (36.0%). It was higher for people with low education and low income compared to those with high education and high income. The sensitivity of the indicator was highest for people residing in Queensland (38.0%) and South Australia (37.9%) and lowest for people in New South Wales (34.4%).

The derived disability indicator has a specificity of 95.9%, meaning that it correctly identified 95.9% of people identified to not have a disability using SDAC.

The specificity of the derived disability indicator was highest for people in younger age groups (above 98% for people aged younger than 25 years; 96% for people aged 25 to 44 years; and between 90 to 94% for people aged 45 years and older). The specificity was notably lower for First Nations Australians compared to non-Indigenous people (91.0% versus 95.5%), for people with low education compared to high education (91.9% versus 97.6%), and for people with low income compared to high income (92.2% versus 96.6%), and similar for men and women and for the different states.

Despite the specificity of the derived disability indicator being high overall and for most subgroups, it is important to acknowledge that a substantial number of people were identified as having a disability by the derived disability indicator only: 1233 individuals, or about 33% of all those identified as having disability by the derived indicator. The specificity remains high because of the large number of people in SDAC who did not have a disability (n=30,061).

Table 14. Sensitivity and specificity of the derived disability indicator against the SDAC disability indicator

		Sensitivity %	Specificity %
Whole sample		36.0	95.9
Age group	0-4	48.9	99.2
	5-14	44.0	98.4
	15-24	42.2	98.5
	25-34	40.4	96.2
	35-44	40.6	96.0
	45-54	44.7	94.1
	55-64	44.9	93.9
	65-74	33.4	90.8
	75+	21.6	93.2
Gender	Men	38.4	95.9
	Women	33.8	95.9
Indigenous status	First Nations Australians	52.3	91.0
	Not Indigenous	36.0	95.6
Education	Completed year 12	31.9	97.6
	Did not complete year 12	38.6	91.9
Income	Highest 60%	21.4	96.6
	Lowest 40%	46.1	92.2
State	NSW	34.4	95.7
	VIC	35.9	96.3
	QLD	38.0	95.8
	SA	37.9	95.6

8.6 Sensitivity analyses

A number of sensitivity analyses were conducted to inform approaches for enhancing the sensitivity and specificity of the derived disability indicators. The first four sensitivity analyses generated new disability indicators, which were tested by repeating the analyses of prevalence and sensitivity and specificity described above. The fifth and sixth sensitivity analyses restricted the age range of the SDAC sample.

Firstly, a quantitative bias analysis was conducted to account for the 5.0% of observations in the AIHW disability cohort that were unable to be linked to MADIP (see Figure 3). We created a new disability variable which included all individuals identified by the derived disability indicator and randomly selected (n=187) additional individuals from those not identified as having a disability by the derived disability indicator in the SDAC sample to account for the 5.0% of observations that were not linked. We randomly selected the observations to maintain the distribution of disability status (from the SDAC disability indicator) that existed for people identified as having a disability by the derived disability indicator in the observed data (67%

disability, 33% no disability), therefore selecting n=126 individuals who had a disability according to the SDAC disability indicator and n=61 individuals who did not have a disability according to the SDAC disability indicator.

Second, we created a modified, more inclusive version of the derived disability indicator. We included additional items from the AIHW linked administrative data to identify individuals with disability who were not included in the original disability indicator. These additional disability items included people who applied for the NDIS but had not received a plan, plus a small number of people who were identified in the SHSC data as having been provided with, referred to or in need of disability services.

Third, we created a disability indicator including people who had been identified as ever having a disability based on the derived overall disability indicator in the AIHW linked administrative data between 2010 and 2020 (not just in the 2018/2019 financial year).

Fourth, we constructed a less inclusive version of the derived disability indicator, restricted to people identified as having a disability in DOMINO, as the analysis of overlap between the datasets illustrated that 98.1% of the cohort was identified using DOMINO (Table 6). The purpose of this was to compare the performance of this less inclusive version with the original derived disability indicator based on disability identification from a variety of administrative data sources.

Fifth, we restricted the sample to people aged 25 to 64 years.

Lastly, we restricted the sample to people aged younger than 65 years.

Table 15. Estimates of prevalence, sensitivity and specificity for the sensitivity analyses

	Prevalence (%)		Sensitivity (%)	Specificity (%)
	Derived disability indicator	SDAC disability indicator		
Main analysis (as above)	9.8	17.7	36.0	95.9
Quantitative bias analysis	10.3	17.7	37.8	95.7
Broader disability definition	10.2	17.7	37.1	98.7
Ever disability	10.5	17.7	37.4	95.3
Restricted to DOMINO only	9.6	17.7	35.4	96.0
Restricted to 25-64 years	10.3	14.1	43.5	95.1
Restricted to <65 years	8.4	12.1	43.6	96.5

The results in Table 15 present the findings of the sensitivity analyses. The first four sensitivity analyses which assessed the effect of using different disability indicators did not materially affect the results.

- The quantitative bias analysis increased the estimate of the prevalence of disability from 9.8% to 10.3% and also increased the sensitivity from 36.0% to 37.8%.
- The broader disability definition increased the prevalence, sensitivity and specificity of the derived disability indicator, which suggests that including people who applied (but did not receive) an NDIS plan increases the validity of the indicator.
- The derived disability indicator including people who had ever been identified as having a disability between 2010 and 2020 increased the sensitivity but decreased the specificity slightly.
- The disability indicator using data from DOMINO only had a slightly lower sensitivity compared to the original derived disability indicator.

The final two sensitivity analyses in which the sample was restricted to people aged 25 to 64 years or less than 65 years had larger impacts. The estimated prevalence of disability in these restricted samples was more similar to the SDAC prevalence (25-64 years: 10.3% versus 14.1%; <65 years: 8.4% versus 12.1%). Compared to the overall sample, the sensitivity was considerably higher (25 to 64 years: 43.5%; <65 years: 43.6%; overall sample: 36.0%), and the specificity was similar, demonstrating that the derived disability indicator performed better in people aged less than 65 years. Interestingly, there was no difference in the sensitivity of the derived disability indicator between the sample restricted to people aged 25 to 64 years and the sample restricted to people aged less than 65 years.

The results of the sensitivity analyses demonstrate that:

- the broader definition of disability (including the additional disability items) should be used to construct disability indicators from administrative data;
- the sensitivity of the derived disability indicator was higher when the sample was restricted to people aged younger than 65 years, suggesting the derived disability indicator performs better for this subgroup.

8.7 Analysis of the disability group indicators using the SDAC sample

Prevalence of each disability group using the SDAC sample

We calculated the prevalence of each disability group using the derived indicators and the SDAC indicators. Prevalence was calculated as the number of people identified as having a disability in each of the disability groups over the total sample size. Survey weights were used to make the estimates representative of the Australian population.

Table 16. Prevalence of each disability group using the derived disability indicator and the SDAC disability indicator

	SDAC sample		SDAC sample <65 years	
	Derived disability Indicator (%)	SDAC disability indicator (%)	Derived disability Indicator (%)	SDAC disability indicator (%)
Disability group				
Sensory or speech[a,b]	0.9	6.1	0.7	2.8
Intellectual or learning[c]	1.6	2.7	1.8	2.7
Physical[d]	6.1	11.0	4.4	6.9
Psychosocial[e]	5.4	4.2	5.2	3.9
ABI[f]	0.3	1.2	0.2	0.9
Other[g]	0.7	7.3	0.8	5.0

[a] see Appendix 3 for mapping of the AIHW disability groups
[b] SDAC sensory disability (sight, hearing, speech)
[c] SDAC intellectual disability (difficulty learning or understanding)
[d] SDAC physical disability (including breathing difficulties, chronic or recurrent pain, incomplete use of limbs and more)
[e] SDAC psychosocial disability (including nervous or emotional conditions, mental illness, memory problems, and social or behavioural difficulties)
[f] SDAC head injury, stroke or acquired brain injury
[g] SDAC 'other' disability (restrictions in everyday activities due to other long-term conditions or ailments)

Looking at the whole SDAC sample, the derived indicators substantially underestimate the prevalence of all disability groups with the exception of psychosocial disability, for which the prevalence is higher using the derived disability indicator (Table 16). 'Sensory or speech', 'ABI', and 'Other' are the groups for which the difference in prevalence estimated by the derived indicator and the SDAC indicator is greatest in relative terms.

For people aged under 65 years, differences in prevalence estimated using the derived and SDAC indicators are slightly smaller, but the derived indicator still substantially underestimates prevalence for all disability groups.

Sensitivity and specificity of the derived indicators of disability groups against the SDAC indicators

We calculated the sensitivity and specificity of the derived indicators of disability for each disability group against the SDAC indicators, for the whole SDAC sample and for the subsample of people aged less than 65 years. Analyses were unweighted as they were not intended to be representative of the Australian population. The estimates of sensitivity and specificity are presented in Table 17.

Table 17. Sensitivity and specificity of each derived disability group indicator against the SDAC disability group indicators

	SDAC sample		SDAC sample <65 years	
	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)
Disability group				
Sensory or speech	8.1	99.6	15.1	99.7
Intellectual or learning	32.6	99.4	39.4	99.3
Physical	31.3	96.9	35.3	97.8
Psychosocial	49.3	96.4	54.9	96.7
ABI	14.9	99.9	18.1	99.9
Other	5.3	99.6	7.9	99.6

Specificity of the derived disability group indicators was uniformly very high (>96%). In the whole SDAC sample, the derived indicator for psychosocial correctly identified 96.4% of individuals who did not have psychosocial disability according to SDAC; for all other disability groups, the specificity of the derived indicator was higher, and greater than 99% for four of the disability groups (Sensory or speech; intellectual or learning; ABI; and other). Estimates of specificity changed minimally when the SDAC sample was limited to people aged younger than 65 years.

Sensitivity of the derived disability group indicators was highest for people with psychosocial disability, with 49.3% of people with psychosocial disability according to SDAC identified by the derived indicator, followed by intellectual or learning (32.6%) and physical disability (31.3%). Sensitivity was lowest for 'other' disabilities, with only 5.3% of those with 'other' disabilities in SDAC identified as such by the derived indicator, and sensory or speech disability, which had a sensitivity of 8.1%. Estimates of sensitivity were somewhat higher when the SDAC sample was limited to people aged younger than 65 years.

8.8 Analysis of prevalence, sensitivity and specificity for the indicator of severe disability

Deriving a revised indicator of severe disability

The original derived indicator of severe disability greatly underestimated the prevalence of severe disability (measured as prevalence of 'severe or profound core activity limitation' in SDAC) (Table 8). This was largely because it did not include any information about disability severity from DOMINO, which was the largest data source contributing to the disability cohort.

We used additional disability items to try to enhance the sensitivity of the derived indicator of severe disability, described in Table 19. These were mostly variables selected from DOMINO, but we also assessed whether other sources could entirely be considered to identify people with severe disability, including NDIS, DS NMDS, MBS and SHSC.

We created a new indicator of severe disability (the 'revised indicator of severe disability') which included individuals identified using the original indicator but also individuals identified using some additional items. We tested each additional item individually by cross tabulating the item against the SDAC indicator of severe disability and calculating the positive predicted value – the proportion of people identified by each item that were identified by the SDAC indicator (see Figure 4). The additional items were selected to be included in the revised indicator of severe disability if their positive predicted value (PPV) was greater than 50% – that is, if more than half of the individuals identified using the indicator were identified by the SDAC severe disability indicator. These included manifest conditions for eligibility for the Disability Support Pension identified in DOMINO, people identified as carees in DOMINO, children attending a special school, people receiving the mobility allowance, and all individuals identified as having a disability in the NDIS and MBS datasets (Table 18).

Table 18. Positive predictive value for each additional severe disability item

Data Source	Disability items	PPV (%)
DOMINO	Identified as having an intellectual disability with IQ<70	51.6
	Identified as permanently blind	>50[a]
	Identified as having a manifest medical condition for eligibility for DSP	60.9
	Capacity to work during partial capacity to work episode: lowest weekly hours	31.5
	Capacity to work during long term temporary reduced capacity: lowest weekly hours	<50[a]
	Identified through the adult caree in the caree medical file	59.7
	Identified through the caree in the concession card database	73.3
	Identified through the caree in the carer payment database	60.3
	Received Disability Support Pension	41.5
	Received sickness allowance payment	<50[a]
	Received business services wage assessment tool payment	<50[a]
	Received mobility allowance	50.0
	Received Youth Disability Supplement Payment	47.3
	Attending a special school	50.0
	Identified as having a terminal illness	<50[a]
DS NMDS	Identified as having a disability in the DS NMDS data	35.0
NDIS	Identified as having a disability in the NDIS data	67.8
SHSC	Identified as having a disability in the SHSC data	<50[a]
MBS	Identified as having a disability in the MBS data	61.7

[a] Numbers too small to present actual percentages

Prevalence of severe disability using the SDAC sample

We calculated the prevalence of severe disability using the original and revised derived indicators of severe disability, which we compared to the SDAC severe disability indicator, for the whole SDAC sample and for the subsample of people aged less than 65 years. Survey weights were used to make the estimates representative of the Australian population.

Table 19. Prevalence of severe disability using the original and revised derived indicators of severe disability and the SDAC indicator of severe disability

	SDAC sample		SDAC sample <65 years	
	Derived disability indicator (%)	SDAC disability indicator (%)	Derived disability indicator (%)	SDAC disability indicator (%)
Derived indicators of severe disability				
Original indicator	0.6	5.3	0.7	3.6
Revised indicator	4.5	5.3	3.8	3.6

The prevalence of severe disability was very low for the original derived indicator of severe disability (0.6%, Table 19). The prevalence estimated using the revised indicator of severe disability was similar to SDAC in the overall SDAC sample (4.5% versus 5.3%) and for people aged less than 65 years (3.8% versus 3.6%).

Sensitivity and specificity of the derived severity indicators against the SDAC indicator

We calculated the sensitivity and specificity of the derived original and revised indicators of severe disability against the SDAC severe disability indicator. Analyses were unweighted as they were not intended to be representative of the Australian population.

Table 20. Sensitivity and specificity of the derived original and revised indicators of severe disability against the SDAC indicator of severe disability

	SDAC sample		SDAC sample <65 years	
	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)
Derived indicators of severe disability				
Original indicator	7.4	99.8	13.2	99.8
Revised indicator	46.3	97.9	54.9	98.2

The sensitivity of the original indicator of severe disability was low (7.4%) and the specificity was very high (99.8%, Table 20). For the revised indicator of severe disability, the sensitivity in the overall SDAC sample was 46.3% and 54.9% in those aged younger than 65 years. The specificity of the revised indicator of severe disability was high for both samples, around 98%.

9. Linked data analysis using the Census

Linkage of the derived disability indicators to the Census in MADIP was used to further analyse the validity of the derived overall disability indicator and the revised indicator of severe disability by testing their performance against the Census disability indicator. It should be noted that the Census disability indicator is aligned conceptually with severe or profound core activity limitation in SDAC.

This section contains a description of the methods used and the results of the analyses, including:

- a description of the linkage between datasets and the resulting sample for analysis;
- the estimated prevalence of disability using the derived overall disability indicator, revised indicator of severe disability, and the Census disability indicator;
- the analysis of the sensitivity and specificity of the derived overall disability indicator and the revised indicator of severe disability against the Census disability indicator.

9.1 Linkage between datasets and sample for analysis

The target Census population for the analysis consisted of all Census participants who were residing in the four states for which we had data from AIHW as part of the NDDA test case: New South Wales, Victoria, Queensland and South Australia (n=19,786,698) [42]. The target Census population was linked to the AIHW disability cohort to further test the performance of the derived disability indicators. This was achieved by linking both the target Census population and the AIHW disability cohort to the MADIP spine.

Of the 17,728,474 individuals residing in New South Wales, Victoria, Queensland and South Australia in the target Census population sample, 95.0% linked to the MADIP population spine – the 'Census analytic sample' – and 5.0% could not be linked to the MADIP population spine. Those who did not link were more likely to be younger, male, have higher education, low income, and have a disability.

The Census analytic sample was then linked to the records of the individuals in the AIHW disability cohort (n=1,783,162) who were identified as having a disability in the 2016/2017 financial year based on the derived disability indicator.

In the Census analytic sample, 98.3% of records were identified in the Medicare Consumer Directory, suggesting that 98.3% of the Census records in MADIP were from the same underlying population as the AIHW disability cohort. Therefore, the difference in the underlying populations is likely to have had only a small impact on the results of the analysis.

9.2 Prevalence of disability in the Census population

We calculated the prevalence of overall disability in the Census using the derived disability indicator, the revised indicator of severe disability, and the Census disability indicator. The analysis was unweighted as the Census aims to cover the whole population.

The Census uses a 'Core Activity Need for Assistance' question module to identify people "needing help or assistance in one or more of the three core activity areas of self-care, mobility and communication, because of a long-term health condition (lasting six months or more), a disability (lasting six months or more), or old age". It is designed to align conceptually with severe or profound core activity limitation in SDAC [16].

Table 21. Prevalence of disability using the derived overall disability indicator, the revised indicator of severe disability and the Census disability indicator

	Census		Census <65 years	
	Derived disability Indicator (%)	Census disability indicator (%)	Derived disability Indicator (%)	Census disability indicator (%)
Overall disability	10.6	5.6	11.6	3.4
Revised severe disability indicator	4.5	5.6	3.6	3.4

Within the whole Census population in MADIP, the estimated prevalence of disability using the derived disability indicator was 10.6%, compared with 5.6% using the Census disability indicator (Table 21). For people aged under 65 years, the difference in estimated prevalence was greater: 11.6% using the derived disability indicator compared to 3.4% using the Census indicator.

The estimated prevalence of severe disability using the revised indicator of severe disability was 4.5%, similar to the prevalence of disability estimated using the Census disability indicator (5.6%). The prevalence estimates were also similar for people aged younger than 65 years: 3.6% using the revised indicator of severe disability and 3.4% using the Census disability indicator.

9.3 Sensitivity and specificity in the Census population

We calculated the sensitivity and specificity of the derived overall disability indicator and the revised indicator of severe disability against the Census disability indicator. The estimates of sensitivity and specificity are presented in Table 22.

Table 22. Sensitivity and specificity of the derived disability indicator and the revised indicator of severe disability against the Census disability indicator

	Census		Census <65 years	
	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)
Overall disability	59.8	92.4	82.0	90.9
Revised severe disability indicator	47.6	98.0	61.1	98.3

Within the whole Census population in MADIP, the derived disability indicator identified nearly 60% of individuals who were identified as having disability by the Census indicator. Of those not identified by the Census indicator, 92.4% were also not identified by the derived disability indicator. For people aged under 65 years, sensitivity was higher, at 82.0%, and specificity was slightly lower, at 90.9%.

The sensitivity of the revised indicator of severe disability against the Census disability indicator was 47.6% and specificity was high (98.0%). For people aged under 65 years, sensitivity was higher, at 61.1% and specificity was also higher, at 98.3%.

Sensitivity of the derived overall disability indicator relative to the Census disability indicator was substantially higher (59.8%) than relative to the SDAC disability indicator (36.0% – Table 13), but specificity was somewhat lower (92.4% against the Census indicator compared with 95.5% against the SDAC indicator). This is as expected, because the Census disability indicator is aligned conceptually with severe or profound core activity limitation in SDAC, and so captures a subgroup of the SDAC disability cohort who experience more severe activity limitations.

10. Linked data analysis of housing outcomes

In order to understand the impact of using the derived disability indicator for monitoring outcomes, we conducted an analysis examining the distribution of housing characteristics between people with and without disability, in the SDAC sample and the Census population.

We examined three housing variables:

- Home ownership (outright owner or mortgage; other)
- Low income private renters, defined as people living in private rental accommodation who were in the lowest 40% of the population income distribution of personal income, (low income private renters; other)
- Living in public housing (renting from a public landlord; other)

In the SDAC sample, we compared housing outcomes for people with and without disability using (1) the derived overall disability indicator, (2) the SDAC disability indicator, (3) the revised indicator of severe disability, and (4) the SDAC severe disability indicator. In the Census population, we used three disability indicators: (1) the derived overall disability indicator, (2) the revised indicator of severe disability, and (3) the Census disability indicator.

For each disability indicator, we calculated the proportion of people with disability and without disability who experienced each of these housing characteristics, using survey weights to make the estimates representative of the Australian population (for the SDAC sample only). We fitted logistic regression models for each housing outcome conditional on disability status, age group and gender to estimate the relative odds of experiencing each outcome for people with disability compared to people without disability.

Table 23. Home ownership for people with and without disability and results of the logistic regression analysis

		No disability (%)	Disability (%)	OR[a,b] (95% CI)
SDAC sample				
Overall disability	Derived disability indicator	63.4	38.2	0.2 (0.2, 0.2)
	SDAC disability indicator	61.7	55.3	0.5 (0.4, 0.5)
Severe disability	Revised indicator of severe disability	61.7	33.4	0.2 (0.2, 0.2)
	SDAC severe disability	61.4	44.6	0.3 (0.3, 0.4)
Census population				
Overall disability	Derived disability indicator	70.2	51.5	0.5 (0.5, 0.5)
Severe disability	Revised indicator of severe disability	68.6	52.9	0.4 (0.4, 0.4)
	Census disability indicator	68.5	57.3	0.3 (0.3, 0.3)

[a] adjusted for age group and gender
[b] compared to no disability

All disability indicators showed that people with disability were less likely to own their own home compared with people without disability (Table 23).

Looking at the odds ratios (adjusted for age group and gender), in the SDAC sample, the odds of home ownership were 80% lower for people with disability relative to people without disability measured using the derived disability indicator, compared to 50% lower using the SDAC disability indicator. For severe disability, the magnitude of the odds ratio was larger for the revised indicator of severe disability compared to the SDAC severe disability indicator.

In the Census population, the odds ratio was of larger magnitude for the Census disability indicator compared to the derived overall disability indicator and the revised indicator of severe disability.

Table 24. Proportion of low-income private renters for people with and without disability and results of the logistic regression analysis

		No disability (%)	Disability (%)	OR[a,b] (95% CI)
SDAC sample				
Overall disability	Derived disability indicator	4.5	11.7	3.4 (2.8, 4.1)
	SDAC disability indicator	4.8	8.1	2.2 (1.9, 2.7)
Severe disability	Revised indicator of severe disability	5.0	14.4	3.8 (2.9, 5.1)
	SDAC severe disability	5.2	9.6	2.4 (1.8, 3.4)
Census population				
Overall disability	Derived disability indicator	6.3	17.9	3.4 (3.3, 3.4)
Severe disability	Revised indicator of severe disability	7.3	18.4	3.4 (3.4, 3.4)
	Census disability indicator	7.3	17.6	4.0 (4.0, 4.0)

[a] adjusted for age group and gender
[b] compared to no disability

All disability indicators showed that people with disability were more likely to be low-income private renters compared with people without disability (Table 24).

In the SDAC sample, the logistic regression analysis demonstrated that the odds of being a low-income renter were more than three times higher for people with disability measured using the derived disability indicator relative to people without disability, compared to 2.2 times higher when using the SDAC disability indicator. For severe disability, the magnitude of the odds ratio was also higher for the revised indicator of severe disability compared to the SDAC severe disability indicator.

In the Census population, the odds ratio was higher for the Census disability indicator (four times higher relative to people without disability), followed by the revised indicator of severe disability and the derived overall disability indicator (OR=3.4).

Table 25. Proportion of people in public housing for people with and without disability and results of the logistic regression analysis

		No disability (%)	Disability (%)	OR[a,b] (95% CI)
SDAC sample				
Overall disability	Derived disability indicator	0.6	9.8	16.3 (12.6, 21.1)
	SDAC disability indicator	1.0	5.4	5.5 (4.3, 7.0)
Severe disability	Revised indicator of severe disability	1.3	11.4	9.8 (7.3, 13.2)
	SDAC severe disability	1.4	8.0	5.7 (4.1, 7.9)
Census population				
Overall disability	Derived disability indicator	1.4	14.4	6.8 (6.8, 6.9)
Severe disability	Revised indicator of severe disability	2.4	18.1	8.7 (8.6, 8.8)
	Census disability indicator	2.4	15.4	11.9 (11.8, 12.0)

[a] adjusted for age group and gender
[b] compared to no disability

Consistent across all disability indicators, people with disability were more likely to be living in public housing compared with people without disability (Table 25).

In the SDAC sample, the odds of living in public housing were 16-times greater for people with disability compared to those without disability, using the derived overall disability indicator. The odds were estimated to be 5.5 times higher using the SDAC disability indicator relative to people without disability. For severe disability, the magnitude of the odds ratio was higher for the revised indicator of severe disability compared to the SDAC severe disability indicator.

In the Census population, the odds of living in public housing were almost 12 times higher using the Census disability indicator, more than 8 times higher for the revised indicator of severe disability, and almost 7 times higher for the derived overall disability indicator.

In summary, for all housing outcomes, in the SDAC sample, the derived disability indicators (both overall and severe) produced larger estimates of inequalities in housing outcomes between people with and without disability compared to the SDAC disability indicators. In the Census population analyses, both the derived overall disability indicator and the revised indicator of severe disability led to smaller estimates of inequalities compared to the Census disability indicator.

11. Discussion of the findings of the linked data analysis

In this section, we highlight the main findings, discuss the limitations of the analysis, and consider the implications for using the derived disability indicators.

The linked data analysis contributes evidence to determine whether the derived indicators accurately identify people with disability and which subgroups of people with disability may be more or less accurately represented in the data. This is essential for understanding the uses of the derived disability indicators and their limitations.

11.1 Key findings

The key findings from the linked data analysis are presented below.

1. **The derived indicator underestimated the prevalence of disability compared to the SDAC disability indicator, particularly for people aged 65 years and older.**

   This finding was consistent with our previous analysis in section 7 in which we compared the prevalence of disability estimated using the derived disability indicator in the AIHW linked data against population statistics, which also found that the prevalence of disability was lower using the derived disability indicator compared to estimates from SDAC (see Section 7.4).

   The majority of the difference in the prevalence is likely to be explained by (1) differences in the operational definitions of disability between SDAC and in the administrative data sources, and (2) differences in the characteristics of people in the SDAC sample and those in the administrative data population, which is based on access to support services and payments (i.e., not all people with disability receive support payments or are NDIS participants, for example).

   In addition, the lower prevalence estimate is in part explained by the 5.0% of the AIHW disability cohort that could not be linked to MADIP. However, the unlinked records are likely to only account for approximately half a percentage point of the difference in the prevalence estimate between the derived disability indicator and the SDAC disability indicator (as estimated by the quantitative bias analysis in Table 15).

2. **Compared to people identified using the SDAC disability indicator, people identified using the derived disability indicator were more likely to be aged 25 to 64 years, men, First Nations Australians and people with low income, and less likely to be aged 65 years and older.**

   These findings highlight the differences in the characteristics of people identified in population surveys compared to administrative data. This is expected given the known characteristics of people who access support services and payments, who have different demographic and socio-economic characteristics compared to the general population. While SDAC is the most detailed and comprehensive source of disability information in Australia, it is not representative of the whole population of people with disability in Australia. The survey does not include people living in very remote areas, discrete Aboriginal and Torres Strait Islander communities, people living in hotels and short-term caravan parks, religious and educational institutions, hostels for the homeless or night shelters, and correctional institutions.

   - The age differences are likely to be explained by eligibility for certain services, e.g., the DSP (<65 years) and the NDIS (less than 65 years at entry into the scheme). Other administrative data sources, such as aged care data, need to be explored to be able to identify people with disability aged 65 years and older.
   - There are also documented gender differences in access to services, for example men are 16% more likely to receive the DSP compared to women [14] and there are very pronounced gender differences in NDIS participation rates [43].
   - SDAC was not developed to collect data for Indigenous Australians specifically and is likely to underestimate First Nations Australians living with disability because the survey did not include people living in very remote areas and discrete Aboriginal and Torres Strait Islander communities.

- Higher levels of socio-economic disadvantage are to be expected in the cohort of people with disability identified in administrative data. In some cases, contact with services is associated with socioeconomic disadvantage (e.g., accessing housing services), and eligibility criteria (e.g., means testing) for some services such as the DSP limit access to people experiencing socioeconomic disadvantage.

3. **Compared to people identified using the SDAC disability indicator, people identified using the derived disability indicator were more likely to have severe disability and psychosocial disability, and less likely to have a sensory or speech disability or ABI.**

   - People with disability identified by the derived overall disability indicator were more likely to have severe disability compared to those identified by the SDAC indicator. This is expected because eligibility criteria for disability services and payments tend to limit the cohort to people experiencing more severe core activity limitations, rather than capturing the broad population of people experiencing limitations in everyday activities identified in SDAC.

   - More research is needed to understand the differences in the disability groups between the derived indicator and the SDAC indicator.

4. **People who were identified as having disability in SDAC but not by the derived disability indicator (the 'false negatives') were less likely to have an intellectual or learning, psychosocial, or 'other' disability, and less likely to have a severe disability compared to all people with disability identified using the SDAC disability indicator.**

5. **People who were not identified as having disability in SDAC but were by the derived disability indicator (the 'false positives') were younger, more likely to be First Nations Australians and residents of New South Wales compared to all people with disability identified using the SDAC disability indicator. They were much more likely to have a psychosocial disability and less likely to have a sensory or speech, 'other' disability, or ABI.**

6. **The sensitivity of the derived disability indicator was low overall (36.0%) – the derived disability indicator correctly identified just over a third of people identified using the SDAC disability indicator.**

   The low sensitivity of the derived disability indicator was expected, given the differences in the underlying operational definitions of disability of the SDAC disability indicator and the derived overall disability indicator, and the differences in the characteristics of the people in the SDAC sample compared to the administrative data, described above.

7. **There were important differences in the sensitivity of the derived overall disability indicator for different subgroups of the population. Sensitivity was higher for people aged less than 65 years, men, First Nations Australians, people with low education, and people with low income.**

   These differences reflect the characteristics of people identified in administrative data compared to the whole population. The subgroups for which the sensitivity is higher are the subgroups of the population that are more likely to access services or supports, and therefore more likely to be represented in the administrative data.

   Interestingly, while the results of the previous analysis comparing the indicators against published population statistics (described in Section 7) suggested that the derived indicator did not adequately identify people aged 0 to 24 years, the sensitivity of the derived disability indicator was higher (>40%) for all age groups younger than 65 years.

8. **The analysis of the sensitivity of the disability group indicators provided evidence that sensitivity was very low for derived indicators of sensory or speech, ABI, and 'other' disabilities (<15%), and somewhat higher for derived indicators of physical (31.3%), intellectual (32.6%), and psychosocial disability (49.3%).**

   The analysis highlighted important differences in the sensitivity and specificity of the indicators of different disability groups. Further analysis is needed to understand which disability groups were identified in each of the administrative data sources and to understand how additional administrative data sources may be able to be used to identify the disability groups that are currently not accurately captured by the derived disability group indicators.

9. **The sensitivity of the revised indicator of severe disability (46.3%) was higher than the overall disability indicator (36.0%) and was further improved by restricting the sample to people aged less than 65 years (54.9%).**

   The revised indicator of severe disability identified approximately half of the people identified by the SDAC severe disability indicator. People in the administrative data sources were more likely to have more severe disability compared to those identified in the SDAC sample, due to eligibility criteria for services and payments, therefore we would expect the sensitivity of the severe disability indicator to be higher than that of the overall disability indicator.

10. **The derived indicator had high specificity overall – it correctly identified 95.9% of people without disability according to the SDAC indicator – and for most subgroups. Specificity was lower (91-92%) for First Nations Australians, people with low education, and people with low income.**

    The high specificity of the indicator, across almost all subgroups of the population, demonstrates the indicator performed well at correctly identifying people without disability in the population. The lower specificity for some population subgroups needs to be explored further and may reflect that some subgroups of people with disability are not correctly identified in survey data.

11. **Despite the specificity of the derived disability indicator being high, it is important to recognise that a about a third of people identified as having a disability by the derived disability indicator were ‘false positives’ – they were not identified as having disability by the SDAC indicator. The specificity remains high because the number of ‘false positives’ (n=1233) was small relative to the number of people in SDAC who did not have a disability (n=30,061) – the denominator for the estimate of specificity.**

12. **The sensitivity (37.1%) and specificity (98.7%) of the derived disability indicator were higher for the more inclusive version of the derived disability indicator (Table 15), which included people who applied for the NDIS but had not received a plan. This suggests that the derived disability indicator should include all NDIS applicants, not just participants.**

13. **In the analysis of the Census data, we found that the sensitivity of the derived disability indicator (overall disability) was relatively high (59.8%), however the specificity was low (92.4%), measured against the Census disability indicator. The sensitivity of the revised indicator of severe disability was also relatively high (47.6%), though not as high as the derived overall disability indicator, and it had higher specificity (98.0%).**

    The results of the analyses of sensitivity are not surprising because the Census disability indicator is designed to align conceptually with severe or profound core activity limitations in SDAC, therefore we would expect the sensitivity of the derived indicators to be higher than in the SDAC analysis. The lower specificity of the overall disability indicator is also expected because it is not limited to identification of people with severe disability, so will capture some individuals with less severe disability who are not captured by the Census indicator.

14. **The analysis of housing outcomes, in the SDAC sample, demonstrated that the derived disability indicators (both overall and severe) produced larger estimates of inequalities between people with and without disability for all housing outcomes compared to the SDAC disability indicators.**

    Though the inequalities were broadly consistent across all the disability indicators, there were differences in the magnitudes of the associations. The derived overall disability indicator estimated greater inequalities in housing outcomes between people with and without disability compared to the SDAC disability indicator. This is not unexpected, because the derived indicator captures a disability cohort biased towards more severe disability compared with the SDAC sample, and inequalities in housing are greater for people with more severe disability. Furthermore, the derived indicator captures a more socio-economically disadvantaged cohort, because it is based on access to support services and payments, some of which are means-tested, so this cohort would be expected to experience greater housing disadvantage compared to the broader population with disability.

15. **In the Census population, the derived overall disability indicator and the revised indicator of severe disability led to smaller estimates of inequalities between people with and without disability for all housing outcomes compared to the Census disability indicator.**

    This suggests that the Census disability indicator may be capturing a cohort of people with more severe disability compared to the revised indicator of severe disability (and the derived overall disability indicator).

11.2 Limitations of the linked data analysis

Comparing the derived disability indicators against established indicators of disability in a linked dataset is an optimal design to assess the validity of the derived indicators and was the key strength of the linked data analysis.

However, there were also limitations. Firstly, the 5% of observations from the AIHW disability cohort that could not be linked to MADIP would have led to an underestimate of the overall prevalence of disability using the derived disability indicator within the SDAC sample and partly account for the low sensitivity of the indicator. We performed a quantitative bias analysis to adjust for the unlinked records, which estimated that the overall prevalence of disability was 10.3% rather than the 9.8% estimated in the observed data, therefore the impact was likely to be small, accounting for about half a percentage point.

The SDAC analyses were conducted within a survey sample population rather than whole-of-population data, leading to uncertainty attached to the estimates because of sampling variation. However, we also conducted analyses using the Census data, which included the whole population for the four states for which we had data.

We have used the SDAC indicator as a ‘gold standard’ in the analyses of sensitivity and specificity. SDAC is designed to measure disability prevalence and provides the most detailed and comprehensive source of data on disability in Australia. However, SDAC is not representative of the whole population of people with disability in Australia because the sampling design does not include people living in very remote areas, discrete Aboriginal and Torres Strait Islander communities, people living in hotels and short-term caravan parks, religious and educational institutions, hostels for the homeless or night shelters, and correctional institutions.

The SDAC approach to identifying disability is broadly aligned with the ICF; the operational definition of disability in SDAC is “any limitation, restriction or impairment which restricts everyday activities and has lasted, or is likely to last, for at least six months”. We would not expect the derived disability indicator to coincide completely with the SDAC indicator, for the following reasons.

- Firstly, the data sources used to create the derived disability indicator capture people with disability who meet the eligibility criteria for particular services and programmes. For example, people identified through disability-related income support payments in DOMINO met eligibility criteria defined with reference to impairments and employment restrictions (not with reference more broadly to limitations in everyday activities, as in the SDAC definition).
- Secondly, eligibility for payments for some items used in DOMINO was restricted to people on low income (e.g., receipt of DSP). Given that the large majority of people in the derived disability indicator cohort were identified using DOMINO, this may explain in part the lower prevalence of disability for people with higher education and higher income using the derived disability indicator compared to SDAC.

Therefore, the derived disability indicator cannot be regarded as aligned with the SDAC operational definition of disability.

In addition, the derived disability indicator used AIHW administrative data for the whole 2018/2019 financial year to identify disability, whereas data were collected for the SDAC Survey between 29 July 2018 and 2 March 2019. Because of the dynamic nature of disability, we would expect changes in people’s disability status to occur throughout the financial year, which may explain some of the discrepancies between the indicators.

The categories of the derived disability group indicator were aligned as closely as possible with the disability groups used for reporting SDAC data: sensory or speech; intellectual or learning; physical; psychosocial; acquired brain injury; other. This approach was appropriate in the context of this test case, to enable us to test the administrative data against a 'gold standard' data source. However, the SDAC groups are based largely on information about functioning limitations, whereas a mix of functioning limitation and diagnostic information available in NDIS, DS NMDS and DOMINO was mapped to the SDAC groups to produce the derived disability group indicator (see Appendix 3). For example, 'autism' (a diagnosis) in NDIS and DS NMDS data was mapped to 'Intellectual disability' in the derived disability group indicator; in SDAC, this group is assigned if a person responds positively to a question about 'difficulty learning or understanding things' (a functioning limitation). This illustrates the limitations of this mapping approach, as autistic people may experience diverse functioning limitations, not necessarily related to difficulty learning or understanding. Further work will be required to refine the categories of the derived disability group indicator for use in the enduring NDDA, informed by the needs of users and information available in different data sources to support the construction of disability group categories.

Similarly, the derived severity indicator will require further refinement. Neither DOMINO or NDIS data allow construction of severity categories that align well with the SDAC severity categories. Future decisions about whether and how a severity indicator should be constructed for use within an enduring NDDA should be informed by consideration of the purpose/s for which this indicator will be used.

Finally, it is important to note that a range of factors can impact the identification of First Nations people with disability in data sources. As explained by members of the Aboriginal and Torres Strait Islander Perspectives Reference Panel for the NDDA Pilot, such factors include whether the language used to ask about disability is culturally appropriate, concerns people may have about disclosing information related to disability, and barriers to accessing services and programs (including because of geographic and cultural factors). Similar factors may impact the identification of people with disability from culturally and linguistically diverse backgrounds. This is an issue that deserves attention to ensure that the NDDA makes an effective contribution to strengthening the evidence base in relation to all people with disability.

11.3 Recommendations

The findings of this analysis provide insight into the characteristics of the cohort of people with disability identified in administrative data and how this cohort compares to the broader populations of people with disability identified in SDAC and the Census. It is important to understand who is identified in the data to understand the strengths and limitations of the derived disability indicator for different uses and applications.

1. **Administrative data from DOMINO, DS NMDS (for historical analyses) and NDIS (including people who applied but were not granted plans) should be used to construct the derived disability indicators.**
2. **The derived disability indicators should not be used to estimate population disability prevalence.**

The derived disability indicators are administratively based indicators that reflect services used and supports received, however not all people with disability receive support payments or are NDIS participants, for example. Furthermore, the derived overall disability indicator is not based on or aligned with a conceptual definition of disability.

3. **Use of the derived disability indicators for specific purposes should be guided by an understanding of the strengths and limitations of the indicators, and results should be reported with relevant caveats attached.**
    - The findings suggest that the derived disability indicators can be used for comparing service use and outcomes between people aged younger than 65 years with and without disability within administrative data sources, but caveats should always be attached to reported data stating that the derived disability indicator cohort does not capture all people with disability. There should be particular consideration of those subgroups who are known to be under-represented – people with mild and moderate disability, older people with disability, women, non-Indigenous people, and those with higher education and income.
    - One of the key contributions of linked data is its use in analyses of rare outcomes, small subgroups of the populations, and fine-grained temporal analysis of events, which cannot be examined in survey

data. The derived disability indicators may be able to be used for these purposes, again stating the caveats that the derived disability indicator cohort does not capture all people with disability.

4. **Further research is needed to refine the derived disability indicators.**

Further research is needed to:

- further refine the derived disability indicators, using additional data sources such as education and aged care data, to ensure better representation of demographic and socio-economic subgroups of the population in the NDDA;
- quantitatively explore the extent to which the differences in the underlying populations account for the differences in the estimates of disability, either by including the whole AIHW population spine (including people without disability) or by identifying the subset of the population in MADIP which corresponds to the population underlying the AIHW disability cohort;
- better understand the discrepancies between the derived disability indicator and the SDAC disability indictor – the 'false positives' and the 'false negatives' – and implications for analyses of outcomes;
- apply analytic techniques that try to 'correct' for biases such as quantitative bias analysis, which may expand the uses of the derived disability indicators for analyses describing the whole population of people with disability in Australia;
- assess whether the derived indicators can be used to examine rare outcomes, small population subgroups, and monitor outcomes over time.

Further research should be led by people with disability and/or conducted by and with people with disability. People with disability should inform and be actively involved in further research into the development and use of derived disability indicators in linked administrative data.

5. **Data development work is needed to improve disability identification in administrative datasets.**

There is a need for strategies to improve identification of individuals with disability in administrative data sets so that derived indicators can be produced that are more consistent with the ICF conceptual model of disability and the SDAC operational definition.

11.4 Conclusions

These results highlight the potential of the derived disability indicators for use in linked administrative data collections. Linked administrative data can provide insights into contacts with the disability and mainstream service system and related life outcomes (e.g., health, education and employment) for people with disability, and disability-related inequalities that cannot be obtained from administrative data.

However, it is important to understand the limitations of the derived disability indicators, which are not representative of the whole population with disability as 'disability' is understood in Australia. While this is a limitation of the data, it is to be expected. Future work should focus on understanding in more detail the population of people with disability administrative data can identify. To do this, it is vital that data on the whole Australian population, where possible, is gradually integrated into the NDDA.

Importantly, people with disability and their representative organisations must be key players in the development of disability data and statistics, and in their use.

12. References

1. Australian Civil Society CRPD Shadow Report Working Group, *Australian Civil Society Shadow Report to the United Nations Committee on the Rights of Persons with Disabilities* 2019.
2. Davy, L., et al., *Review of implementation of the National Disability Strategy 2010-2020: Final report. (SPRC Report [4/19)*. 2019, Social Policy Research Centre, UNSW Sydney: Sydney.
3. Productivity Commission, *Disability care and support. Productivity commission inquiry report. No. 54*. 2011: Canberra.
4. World Health Organization and United Nations ESCAP, *Training manual on disability statistics*. 2008, United Nations: Bangkok.
5. World Health Organization. *How to use the ICF: A practical manual for using the International Classification of Functioning, Disability and Health (ICF). Exposure draft for comment*. 2013; Available from: http://www.who.int/entity/classifications/drafticfpracticalmanual2.pdf?ua=1.
6. World Health Organization, *International Classification of Functioning, Disability and Health*. 2001, Geneva (CH): WHO.
7. World Health Organization and World Bank, *World report on disability*. 2011, WHO: Geneva (CH).
8. Madden, R.H. and A. Bundy, *The ICF has made a difference to functioning and disability measurement and statistics.* Disability and Rehabilitation, 2018. **41**(12): p. 1450-1462.
9. Madden, R. and R. Madden, *Disability services and statistics: past, present and future*, in *Australia's welfare 2019 data insights*, Australian Institute of Health and Welfare, Editor. 2019, AIHW: Canberra (AU).
10. Australian Institute of Health and Welfare, *Current and future demand for specialist disability services. Disability series. Cat. no. DIS 50*. 2007, AIHW: Canberra.
11. Australian Bureau of Statistics, *4431.0.55.002 - ABS Sources of Disability Information, 2012 - 2016*. 2018, ABS: Canberra (AU).
12. Australian Bureau of Statistics, *4430.0 - Disability, Ageing and Carers, Australia: Summary of Findings, 2018*. 2019, ABS: Canberra (AU).
13. Australian Bureau of Statistics, *Disability, Ageing and Carers, Australia: Summary of Findings methodology*. 2019, ABS: Canberra.
14. Australian Institute of Health and Welfare, *People with disability in Australia 2020*. 2020, AIHW: Canberra.
15. Productivity Commission. *Report on Government Services 2018*. 2018 3 August 2020]; Available from: https://www.pc.gov.au/research/ongoing/report-on-government-services/2018.
16. Australian Bureau of Statistics, *2901.0 - Census of Population and Housing: Census Dictionary, 2016*. 2016, ABS: Canberra (AU).
17. Australian Institute of Health and Welfare, *Standardised Disability Flag for mainstream services: data collection guide. Cat. no. DAT 6*. 2016, AIHW: Canberra.
18. Balogh, R., et al., *Data Linkage: Canadian and Australian Perspectives on a Valuable Methodology for Intellectual and Developmental Disability Research.* Intellect Dev Disabil, 2019. **57**(5): p. 439-462.
19. Florio, T. and J. Trollor, *Mortality among a Cohort of Persons with an Intellectual Disability in New South Wales, Australia.* J Appl Res Intellect Disabil, 2015. **28**(5): p. 383-93.
20. Hwang, Y.I.J., et al., *Mortality and cause of death of Australians on the autism spectrum.* Autism Res, 2019. **12**(5): p. 806-815.
21. Reppermund, S., et al., *Factors associated with death in people with intellectual disability.* J Appl Res Intellect Disabil, 2020. **33**(3): p. 420-429.
22. Reppermund, S., et al., *Cohort profile: a data linkage cohort to examine health service profiles of people with intellectual disability in New South Wales, Australia.* BMJ Open, 2017. **7**(4): p. e015627.
23. Reppermund, S., et al., *Health and wellbeing of people with intellectual disability in New South Wales, Australia: a data linkage cohort.* BMJ Open, 2019. **9**(9): p. e031624.

24. Trofimovs, J., et al., *Using linked administrative data to determine the prevalence of intellectual disability in adult prison in New South Wales, Australia.* J Intellect Disabil Res, 2021. **65**(6): p. 589-600.
25. Balogh, R., et al., *Hospitalisation rates for ambulatory care sensitive conditions for persons with and without an intellectual disability--a population perspective.* J Intellect Disabil Res, 2010. **54**(9): p. 820-32.
26. Lin, E., et al., *Using administrative health data to identify individuals with intellectual and developmental disabilities: a comparison of algorithms.* J Intellect Disabil Res, 2013. **57**(5): p. 462-77.
27. Ben-Shalom, Y. and D.C. Stapleton, *Predicting Disability among Community-Dwelling Medicare Beneficiaries Using Claims-Based Indicators.* Health Serv Res, 2016. **51**(1): p. 262-81.
28. Clements, K.M., et al., *Pregnancy Characteristics and Outcomes among Women at Risk for Disability from Health Conditions Identified in Medical Claims.* Womens Health Issues, 2016. **26**(5): p. 504-10.
29. Darney, B.G., et al., *Primary Cesarean Delivery Patterns among Women with Physical, Sensory, or Intellectual Disabilities. Time trends in births and cesarean deliveries among women with disabilities.* Womens Health Issues, 2017. **27**(3): p. 336-344.
30. Horner-Johnson, W., et al., *Time trends in births and cesarean deliveries among women with disabilities.* Disability and Health Journal, 2017. **10**(3): p. 376-381.
31. Khoury, A.J., et al., *The association between chronic disease and physical disability among female Medicaid beneficiaries 18-64 years of age.* Disabil Health J, 2013. **6**(2): p. 141-8.
32. Palsbo, S.E., et al., *Identifying and classifying people with disabilities using claims data: further development of the Access Risk Classification System (ARCS) algorithm.* Disabil Health J, 2008. **1**(4): p. 215-23.
33. Van Naarden Braun, K., et al., *Trends in the prevalence of autism spectrum disorder, cerebral palsy, hearing loss, intellectual disability, and vision impairment, metropolitan atlanta, 1991-2010.* PLoS One, 2015. **10**(4): p. e0124120.
34. Emerson, E., *Deprivation, ethnicity and the prevalence of intellectual and developmental disabilities.* J Epidemiol Community Health, 2012. **66**(3): p. 218-24.
35. Emerson, E., Hatton, C., Robertson, J., Roberts, H., Baines, S., Evison, F.,& Glover, G, *People with learning disabilities in England 2011. Public Health England report*, in *Improving Health and Lives (IHAL)*. 2012.
36. Emerson, E. and G. Glover, *The “transition cliff” in the administrative prevalence of learning disabilities in England.* Tizard Learning Disability Review, 2012. **17**(3): p. 139-143.
37. Kiani, R., et al., *Urban-rural differences in the nature and prevalence of mental ill-health in adults with intellectual disabilities.* J Intellect Disabil Res, 2013. **57**(2): p. 119-27.
38. Brown, H.K., et al., *Identifying reproductive-aged women with physical and sensory disabilities in administrative health data: A systematic review.* Disabil Health J, 2020. **13**(3): p. 100909.
39. Friedman, D.J., R.G. Parrish, and M.H. Fox, *A Review of Global Literature on Using Administrative Data to Estimate Prevalence of Intellectual and Developmental Disabilities.* J Policy Pract Intellect Disabil, 2018. **15**(1): p. 43-62.
40. Australian Bureau of Statistics, *44300DO090_2018 Survey of Disability, Ageing and Carers: Aboriginal and Torres Strait Islander People with Disability*. 2021, ABS: Canberra (AU).
41. Australian Institute of Health and Welfare, *National Disability Data Asset: Outcomes Reporting – Linkage Report (EO2020-3-1185)*. 2021, AIHW: Canberra.
42. Australian Bureau of Statistics, *2071.0 - Census of Population and Housing: Reflecting Australia - Stories from the Census, 2016 - Snapshot of Australia*. 2017, ABS: Canberra (AU).
43. Yates, S., et al., *Women's experiences of accessing individualized disability supports: gender inequality and Australia's National Disability Insurance Scheme.* Int J Equity Health, 2021. **20**(1): p. 243.

Appendix 1. Studies included in the literature review

The table below summarises, for each of the studies included in the literature review, the target population, the datasets and disability items used to identify disability, the derivation of the disability indicator, and any validity testing.

Author (year) state or country	Disability population	Datasets used for disability identification	Disability identification items	Derivation of disability indicator	Methods and results of validity testing
AUSTRALIAN STUDIES					
Florio and Trollor (2015) NSW **[19]**	Intellectual disability	• Ageing, Disability and Home Care (ADHC)	Inclusion in the NSW ADHC intellectual disability service register	Single disability indicator	No validation, however the results suggest that people with mild intellectual disability are likely to be underrepresented.
Reppermund *et al.* (2017) NSW **[22]**	Intellectual disability	• Disability Services Minimum Dataset (DS MDS) Admitted Patients Data Collection (APDC) • Emergency Department Data Collection (EDDC)	**DS-MDS:** People recorded as having intellectual disability as primary or secondary disability **APDC, EDDC:** ICD-10 codes (F700-F701;F708-F709; F710-F711; F718-F719; F720-F721; F728-F729;F730-F731; F728-F729; F730- F731; F738-F739; F780-F781;F788-F791; F843-F844; F798-F799; Q900-Q902; Q909;Q910-Q912; Q913; Q914-Q916; Q917; Q930-Q939; Q992;P043; Q860; Q861; Q862; Q868 Q870-Q873; Q875;Q878; Q898)	Inclusive disability indicator from linked datasets	No validation, however the results suggest that people with mild intellectual disability are likely to be underrepresented.
Balogh *et al.* (2019) WA **[18]**	Intellectual disability	WA Intellectual Disability Exploring Answers database (IDEA): • Disability Services Commission (DSC) client register • Department of Education	**DSC**: meets criteria for intellectual disability (determined by psychometric and adaptive behaviour assessments) **Education**: identified for educational support using psychometric and adaptive behaviour assessments	Inclusive disability indicator from linked datasets	No validation. The results demonstrate that adding education data to DSC better identified people with intellectual disability. Linked DSC and education data identified intellectual disability better than hospital data (20% missed using ICD codes only).

Hwang *et al.* (2019) NSW **[20]**	Intellectual disability and autism Age 5-64 years	**Service datasets** • Disability Services Minimum Dataset (DS MDS) • NSW Department of Education • State-wide Disability Service • NSW Public Guardian and NSW Ombudsman **Health datasets** • NSW Admitted Patient Data Collection • Mental Health Ambulatory Data Collection • NSW Emergency Department Data Collection	**Service datasets:** People recorded as having autism or intellectual disability as primary or secondary disability **Health datasets** ICD-10 codes (F84.0, F84.1, F84.5, F84.8, F84.9) used identify people with Autism; ICD-10 codes (specific ICD codes not listed in paper) used to identify people with intellectual disability	Inclusive disability indicator from linked datasets	No validation.
Reppermund *et al.* (2019) NSW **[23]**	Intellectual disability	• Disability Services Minimum Dataset (DS MDS) • Admitted Patients Data Collection (APDC) • Emergency Department Data Collection (EDDC) • Mental Health Ambulatory Data Collection (MHADC) • Corrective Services • State-wide Disability Services • Public Guardian • Disability in Public Schools Dataset (D-PS) • NSW Ombudsman	**DS-MDS:** People recorded as having intellectual disability as primary or secondary disability **APDC, EDDC:** ICD-10 codes **MHADC:** DSM-IV **D-PS**: Receiving specialist support services in public schools **Other**: intellectual disability flag (not further described in article)	Inclusive disability indicator from linked datasets	No validation, however the results suggest that people with mild intellectual disability are likely to be underrepresented.
Reppermund *et al.* (2020) NSW **[21]**	Intellectual disability	• Disability Services Minimum Dataset	People recorded as having intellectual disability as primary or secondary disability (2005-2015)	Single disability indicator	No validation, however the results suggest that people with mild intellectual disability are likely to be underrepresented.
Trofimovs *et al.* (2021) NSW **[24]**	Intellectual disability	• Disability Services Minimum Dataset (DS MDS) • Admitted Patients Data Collection (APDC) • Mental Health Ambulatory Data Collection (MHADC) • Public Guardian • Corrective Services	**DS-MDS:** People recorded as having intellectual disability as primary or secondary disability **APDC:** ICD-10 codes **MHADC, Public Guardian:** DSM-IV **Corrective services:** Predominantly using the Wechsler	Inclusive disability indicator from linked datasets	No validation, however the estimated prevalence of intellectual disability of 4.3% among adult custody population is likely to be underestimated.

			Adult Intelligence Scale Revised measure of intelligence		

INTERNATIONAL STUDIES

Palsbo *et al.* (2008)* United States **[32]**	All disability types	• Health insurance claims	ICD-9 codes, procedural codes, need for assistance accessing care	Access Risk Classification System (ARCS) algorithm which combines information on health conditions, health system accommodations, and care coordination needs	**Validation**: against self-reported disability from a linked survey, demonstrating that: • diagnosis data alone do not adequately predict functioning; • poor identification of mild disability and better identification of more severe disability; • better identification of disability when multiple data sources were used.
Balogh *et al.* (2010) Canada **[25]**	Intellectual disability	• Medical services • Hospital records • Family services • Education enrolment	**Medical**: ICD-9 codes **Family services**: receiving income assistance for intellectual disability **Education**: special educational data for 'multiple handicaps'	Inclusive disability indicator from linked datasets	No validation, however the results suggest that people with mild intellectual disability are likely to be underrepresented.
Lin *et al.* (2012) Canada **[26]**	Intellectual disability	• Ontario Mental Health Reporting System • Discharge Abstract Database • Same Day Surgery Database • National Ambulatory Care Reporting System • Ontario Health Insurance Plan	ICD-9, ICD-10 and DSM-4 codes	3 algorithms generated based on frequency of contact with health services.	No validation, however the results suggest that using health data alone underestimates the prevalence of intellectual disability.
Emerson and Glover (2012) United Kingdom **[36]**	Learning disability	• Education data (children) • Medical data (GP practice)	**Education**: Special educational needs assessment **Medical**: Included in GP learning disability register	Inclusive disability indicator from linked datasets	No validation.
Emerson (2012) United Kingdom **[34]**	Intellectual and developmental disability	• Education data (School Census)	Special educational needs assessment	Single disability indicator	No validation, however the estimated prevalence rates of intellectual disability and autism

	Age 7-15 years				are consistent with results of previous research.
Emerson *et al.* (2012) United Kingdom **[35]**	Learning disability	• Medical data (GP practice) • Education data • Social care services	**Medical**: Included in GP learning disability register **Education**: Special educational needs assessment, Learning disability register **Social care services**: not described	Disability items from each dataset examined individually	No validation.
Kiani *et al.* (2013) United Kingdom **[37]**	Intellectual disability	• Intellectual disability service register	Assessed using the Leicestershire Intellectual Disability Scale, which combines questions on intelligence, adaptive behaviour and dependency	Single disability indicator	No validation, however the results suggest that people with mild intellectual disability are likely to be underrepresented.
Khoury *et al.* (2013) United States **[31]**	Physical disability	• Health insurance eligibility and claims data	ICD-9 codes, procedural codes	ICD-9 codes associated with physically-disabling conditions and procedural codes for mobility-assistive devices to define three levels of physical disability (none; disability but no mobility device; disability and mobility device)	No validation.
Van Naarden Braun *et al.* (2015) United States **[33]**	Developmental disability (children only)	• Developmental disabilities surveillance program • Education data • Health data	**Medical**: ICD-9 codes and DSM-4 codes related to intellectual disability, autism, cerebral palsy, hearing loss and vision impairment **Education**: special education eligibility.	Single disability indicator	No validation, however the results suggest that people with mild intellectual disability are likely to be underrepresented.
Clements *et al.* (2016) United States **[28]**	All disability types	• Hospital discharge data	ICD-9 codes, procedural codes, prescriptions, need for assistance accessing care	Access Risk Classification System (ARCS) algorithm modified to exclude prescriptions and diagnoses of non-malignant neoplasms	No validation.

Ben-Shalom *et al.* (2016) United States **[27]**	All disability types	• Health insurance claims	ICD-9 codes, procedural codes, prescriptions, need for assistance accessing care, Chronic Illness and Disability Payment System (CDPS)	Constructed six claims-based disability indicators: • Access Risk Classification System (ARCS) highest-risk class; • CDPS 'medium cost'; • SSA-HIT listings of impairments from ICD-9 and procedural codes; • ICD-9 codes for (1) psychiatric disorders, (2) cognitive disorders and (3) intellectual disability.	**Validation**: against self-reported disability from a linked survey, demonstrating that: • Better identification of disability when multiple data sources were used; • Data on service use and prescriptions may improve disability identification.
Darney *et al.* (2017) United States **[29]**	Physical, sensory or intellectual disability	• Hospital discharge data	ICD-9 codes, procedural codes	Khoury method revised to include additional codes relating to physical impairments, and codes relating to sensory and intellectual disability	No validation.
Horner-Johnson *et al.* (2017) United States **[30]**	Physical, sensory or intellectual disability	• Hospital discharge data	ICD-9 codes, procedural codes	Khoury method revised to include additional codes relating to physical impairments, and codes relating to sensory and intellectual disability	No validation.

** this study from 2008 was included as it was the basis for future relevant studies*

Appendix 2. Results of the metadata analysis

Data source	Disability identification items	Information on disability group, severity, temporality
DS NMDS **Population:** National Disability Agreement (NDA) service users **Age range:** 95% <65 years **Temporality:** Annual (by financial year); Start-end (for individual service type outlets: service start date; date service last received; service exit date) **Dataset specification:** https://meteor.aihw.gov.au/content/index.phtml/itemId/698074 **Data quality statement**: https://meteor.aihw.gov.au/content/index.phtml/itemId/728139	**SDAC alignment:** Broadly aligned, i.e., focus on concept of activity limitation and need for assistance. Likely to include mainly people with 'severe or profound core activity limitation', so narrower than SDAC all disability. ICF components: AL, EF (assistance/supervision). **Coverage:** • Current DS service user *Note:* In 2014-15 there were 217,000 services users – c.f., 715,000 people with severe or profound disability (30% of) 2,494,000 people with disability (9% of) aged under 65 in 2015. **Additional notes:** • Missing data for some service outlets. • Service user data not reported for some service types. • Records with missing or invalid statistical linkage keys (0.4%) treated as belonging to separate individuals (possible overestimation of number of services users). • Coverage/completeness varies between jurisdictions. • DS NMDS now discontinued. Transition to NDIS may have had some data quality impacts over transition period.	**Disability group**: *Primary disability group* *Other significant disability group(s)* (Categories: 1 Intellectual (including Down syndrome); 2 Specific learning/Attention Deficit Disorder (other than Intellectual); 3 Autism (including Asperger's syndrome and Pervasive Developmental Delay); 4 Physical; 5 Acquired brain injury; 6 Neurological (including epilepsy and Alzheimer's disease); 7 Deafblind (dual sensory); 8 Vision; 9 Hearing; 10 Speech; 11 Psychiatric; 12 Developmental delay) For primary disability, if discrepancies between data from different agencies exist, records from services with high frequency of contact (e.g., accommodation services) are deemed more accurate. The disability group from the less accurate source will be included as a secondary disability. **Severity**: *Frequency of support required for life activity* – for 9 life areas, including self-care, mobility and communication (can be used to produce 'severe or profound core activity limitation' cohort) **Temporality of disability information:** Information about disability is reported for each service user by financial year. Some states supply data annually while others quarterly – the most recent data in a financial year is used in linked data source.
NDIS **Population:** NDIS participants **Age range:** <65 years at scheme entry **Temporality:** Start-end (Date of the access request decision; Date when first plan was approved; Month of exit)	**SDAC alignment:** NDIS targets people with 'permanent and significant disability', i.e., the disability is likely to be lifelong and has a substantial impact on ability to complete everyday activities. While not limited to people who need assistance, the cohort is likely to be weighted towards people with 'severe or profound core activity limitation', so narrower than SDAC all disability. ICF components: AL	**Disability group**: *Primary disability of the participant* –17 categories (ABI, autism, cerebral palsy, developmental delay, Down syndrome, global developmental delay, hearing impairment, intellectual disability, MS, other, other neurological, other physical, other sensory/speech,

Data source	Disability identification items	Information on disability group, severity, temporality
	Coverage: - Ever NDIS participant *Note:* Estimated to reach 500,000 NDIS participants – c.f. 726,000 people with severe or profound disability (69% of), 2,427,000 people with disability (21% of) aged under 65 in 2018. (Note only 184,000 participants in 2018.)	psychosocial, spinal cord injury, stroke, visual impairment) *Secondary disabilities of the participant* – 77 ICD codes (for information about conversion from ICD codes to the 17 disability categories see: http://.data.ndis.gov.au/data-downloads file name: Participant count by diagnosis data **Severity**: *The participant's normalised severity score* – 15 numbered categories (no description). Advice from data custodian that categories 11-15 likely to correspond to 'severe or profound core activity limitation' **Temporality of disability information:** disability group and severity Information is most recent available for each NDIS participant (i.e., no record of disability group or severity for previous years available in linked dataset).
DOMINO ***Disability Support Pension (DSP)*** (BEN_TYPE_CODE = DSP) **Population:** Total population **Age range:** >15 years, <age pension age **Temporality:** Start-end (Event End Date; Event Start Date) **Eligibility criteria:** https://www.servicesaustralia.gov.au/individuals/services/centrelink/disability-support-pension/who-can-get-it	**SDAC alignment:** Impairment and participation restriction (inability to work independently within the next 2 years). Does not align with temporal dimension of SDAC disability definition ('at least 6 months'). ICF components: BF, PR **Coverage:** - additional criteria restrict eligibility (incl. income and assets tests, participation requirements) *Note:* 23% of people with disability living in households and 47% of those with severe or profound disability were receiving DSP (2018 SDAC)	**Disability group**: *Primary Medical Code; Secondary Medical Code*: 23 categories (ACQUIRED BRAIN IMPAIRMENT, AMPUTATION, CANCER/TUMOUR, etc) *Note*: condition recorded as *Primary Medical Code* only if 'permanent' or 'fully diagnosed, treated and stabilised' See also *Secondary Medical Code Permanent Indicator* – Missing, PRM DTS, PRM NONDTS, PERMANENT, RECURRING, TEMPORARY **Severity**: potentially relevant information from items *Impairment Rating, Manifest condition, Incapacity* **Temporality of disability information**: disability information is current, updated at each assessment.
DOMINO ***Youth Disability Supplement*** CMPNT_TYPE=YDS or YDSO **Population:** Total population	**SDAC alignment:** Illness, injury or disability that both stops you from working 30 or more hours a week and will last more than 2 years; i.e., Disability/health condition + participation restriction (work). Does not	As for DOMINO – *Disability Support Pension (DSP)*

Data source	Disability identification items	Information on disability group, severity, temporality
Age range: <15 years, <22 years **Temporality:** Start-end (Event End Date; Event Start Date) **Eligibility criteria:** https://www.servicesaustralia.gov.au/individuals/services/centrelink/youth-disability-supplement	align with temporal dimension of SDAC disability definition ('at least 6 months') ICF components: HC, BF, PR **Coverage:** • people who receive DSP, Youth allowance, or ABSTUDY	
DOMINO ***Sickness Allowance*** BEN_TYPE_CODE = SKA **Population:** Total population **Age range:** >21/>24, <66/67 years **Temporality:** Start-end (Event End Date; Event Start Date) **Eligibility criteria:** http://www.chronicillness.org.au/workwelfarewills/centrelink-entitlements/sickness-allowance/	**SDAC alignment:** Inability to work due to a temporary illness or accident. Does not align with temporal dimension of SDAC disability definition ('at least 6 months') ICF components: HC, PR **Coverage:** • eligibility criteria incl. income and assets tests, medical certificate confirming inability to work/study **Additional notes:** Sickness Allowance was closed to new recipients on 20 March 2020 and ceased completely on 20 September 2020.	As for DOMINO – *Disability Support Pension (DSP)*
DOMINO ***Mobility Allowance*** BEN_TYPE= MOB **Population:** Total population **Age range:** >16 years **Temporality:** Start-end (Event End Date; Event Start Date)	**SDAC alignment:** A payment to help with travel costs for work, study or looking for work if you have a disability, illness or injury that means you can't use public transport. i.e., Disability/health condition + activity limitation (mobility). ICF components: PR **Coverage:** • eligibility requires need to travel for study, training, work or to look for work. **Additional notes:** People with disability cannot receive the mobility allowance once they have an NDIS plan.	As for DOMINO – *Disability Support Pension (DSP)*
DOMINO ***Medical Equipment Payment*** BEN_TYPE= MEP CMPNT_TYPE=EMEP **Population:** Total population **Age range:** not specified	**SDAC alignment:** A yearly payment to help with energy costs to run medical equipment or medically required heating or cooling (See list of eligible medical conditions). Poor SDAC alignment (no activity limitation/participation restriction specified) ICF components: HC **Coverage:**	As for DOMINO – *Disability Support Pension (DSP)*

Data source	Disability identification items	Information on disability group, severity, temporality
Temporality: Start-end (Event End Date; Event Start Date)	- eligibility criteria include hold a Commonwealth Concession Card, pay for the energy running costs. **Additional notes**: may not be able to determine whether benefit is for medical equipment for person receiving the payment or person they care for.	
DOMINO ***Business Services Wage Assessment Tool (BSWAT) Payment*** BEN_TYPE=BSW **Population:** Total population **Age range:** not specified **Temporality:** Start-end (Event End Date; Event Start Date)	**SDAC alignment:** BSWAT scheme delivered a one-off payment of $100 or more to offer compensation to eligible supported employees with intellectual impairment whose wages were assessed and paid using the BSWAT. Scheme closed 31 Dec 2018. Can assume employment restriction. ICF components: BF, PR **Coverage:** - eligible supported employees with intellectual impairment	**Disability group**: Intellectual disability **Severity**: No information **Temporality of disability information**: unclear, however, may be reasonable to assume intellectual disability is permanent
DOMINO ***Health Care Card child of an adult who received Carer Allowance for that child*** CDA_CHILD **Population:** Total population **Age range:** <16 years **Temporality:** Start-end (Event End Date; Event Start Date)	**SDAC alignment:** Child with disability or severe illness. Poor SDAC alignment (no activity limitation/participation restriction specified). ICF components: unclear **Coverage:** - see eligibility criteria for carer allowance	**Disability group**: No information **Severity**: No information **Temporality of disability information**: assume disability current for duration of Health Care Card child status
DOMINO ***Person attending a special school*** INST_TYPE=O **Population:** Total population **Age range:** children **Temporality:** Start-end (Event End Date; Event Start Date)	**SDAC alignment:** ICF components: PR, EF **Coverage:** - children attending special schools	**Disability group**: No information **Severity**: No information **Temporality of disability information**: assume disability current for duration of special school attendance

Data source	Disability identification items	Information on disability group, severity, temporality
DOMINO ***Primary Medical Code*** MED_PRMY **Population:** Total population **Age range:** not specified **Temporality:** Start-end	**SDAC alignment:** ICF components: HC, BF/BS **Coverage:** • payment applicants who undergo an assessment and have a condition deemed to be 'permanent' or 'fully diagnosed, treated and stabilised'	As for DOMINO – *Disability Support Pension (DSP)*
MBS ITEM = 135, 137, 139, 289, 82000, 82005, 82010, 82030, 82015, 82020, 82025, 82035 • **Group M10** - Autism, Pervasive Developmental Disorder and Disability Services (82000, 82005, 82010, 82030, 82015, 82020, 82025, 82035) • **Group A29** - Early Intervention Services For Children With Autism, Pervasive Developmental Disorder Or Disability (135, 137, 139) • **Group A8** – consultant physician (psychiatry) for assessment, diagnosis and preparation of a treatment and management plan for a patient with autism or another pervasive developmental disorder (item 289). **Population:** Total population **Age range:** <13 years (135, 137, 139, 289, 82000, 82005, 82010, 82030, 82015), <15 years (82020, 82025, 82035) **Temporality:** Event (Date of service)	**SDAC alignment:** These MBS items are based on presence of a diagnosed health condition. Poor SDAC alignment (no activity limitation/participation restriction specified). It may be reasonable to assume that most children identified by these codes will have some degree of activity limitation or participation restriction. ICF components: HC **Coverage:** • Patients who access relevant services from health provider **Additional notes**: • Use of these MBS codes indicates services delivered to an individual for their diagnosed health condition; some are restricted to once-only use for a given individual. Relevant to consider proportion and demographic profile of children with these conditions who ever receive these MBS services. • Further information needed re whether all these MBS codes can only be used where child has relevant diagnosis (e.g., 82030 may be used where 'the child is referred by an eligible practitioner for the purpose of assisting the practitioner with their diagnosis of the child', suggesting that child may not have a diagnosis when this service is provided).	**Disability group**: Group A8 (item 289) is specific to autism and pervasive development disorder, and therefore can identify children with intellectual and learning disability. Items in groups M10 and A29 mention 'Disability' as well as Autism and Pervasive Developmental Disorder, so assume individuals identified by these codes cannot be allocated to a particular disability group on this basis. **Severity**: No information **Temporality of disability information**: identification current at event date
MBS ITEM = 701, 703, 705, 707, 224, 225, 226, 227 • **AN.0.36** Health Assessments (Items 701, 703, 705, 707) • **AN.7.5** Medical Practitioner Health Assessments (Items 224 to 227)	**SDAC alignment:** These MBS items relate to health assessments for a broad range of different health conditions. Poor SDAC alignment (no activity limitation/participation restriction specified) ICF components: HC **Coverage:** Patients who access relevant services from health provider	**Disability group**: No information **Severity**: No information **Temporality of disability information**: N/A

Data source	Disability identification items	Information on disability group, severity, temporality
Population: Total population **Age range:** All **Temporality:** Event (Date of service)	**Additional notes:** These items are not specific to people with disability. People with intellectual disability are among eligible patients for these items, but also various other patient groups; cannot be used to identify patients with disability.	
PBS • **N05A** Antipsychotics • **N05B** Anxiolytics • **N05C** Hypnotics and sedatives • **N06A** Antidepressants • **N06B** Psychostimulants **Population:** Total population **Age range:** All **Temporality:** Event (Date of supply)	**SDAC alignment:** These PBS items relate to prescription of nervous system drugs for diagnosed health conditions. Poor SDAC alignment (no activity limitation/participation restriction specified) ICF components: HC **Coverage:** Patients who receive relevant medication prescription from health provider **Additional notes**: Prescriptions of anxiolytics (N05B), hypnotics and sedatives (N05C), antidepressants (N06A) and psychostimulants (N06B) are used for a wide variety of health conditions, only a fraction of which would be associated with functional limitations. Therefore these items are unlikely to identify people with disability. Prescription of antipsychotics and lithium (N05A) may identify people with psychosocial disability.	**Disability group:** NO5A drugs could be used to identify individuals with psychosocial disability. **Severity:** No information **Temporality of disability information:** Disability identification and disability group current at event date
SHSC Disability requiring assistance flag **Population:** SHSC service users **Age range:** All **Temporality:** Start-end (month) (episode_start_month; episode_start_year; support_period_length; ongoing_support_period_ind) Data quality statement https://meteor.aihw.gov.au/content/index.phtml/itemId/723399	**SDAC alignment:** Identifies people with difficulty, who need assistance, or who use aids/equipment with daily activities of self-care, mobility or communication. Response category '1' Always/sometimes needs help or supervision corresponds to SDAC severe or profound core activity limitation. ICF components: AL+EF **Coverage:** Homelessness support service users	**Disability group**: No information **Severity**: Categories corresponding to SDAC profound, severe, moderate and mild core activity limitation. **Temporality of disability information**: Current at start of support period
PH and SOMIH Disability indicator [dis_p] **Population:** PH/SOMIH service users	**SDAC alignment:** Classifications used for disability status are not consistent across the states and territories, however, states and territories map these data to an AIHW standard. Disability status is derived using the receipt of a disability pension as a proxy in Victoria.	**Disability group**: No information **Severity**: No information **Temporality of disability information**: Current at start of tenancy

Data source	Disability identification items	Information on disability group, severity, temporality
No disability data from VIC – ethics approval not granted No SOMIH data from NSW at this stage – pending ethics approval **Age range:** All **Temporality:** Tenancy start-end (month and year)	New South Wales, Queensland, Tasmania and the Australian Capital Territory reference payment type as well as other information. ICF components: HC + PR (if based on DSP receipt) **Coverage:** Housing support service users. The data cover QLD (PH and SOMIH), SA (PH and SOMIH), NSW (PH only), VIC (none).	

Appendix 3. Mapping of NDIS, DS NMDS and DOMINO disability groups

Disability group	NDIS	DS NMDS	DOMINO
Sensory or speech	Hearing impairment, visual impairment, other sensory or speech	Vison, hearing, deafblind, speech	Sense organs
Intellectual or learning	Intellectual disability, autism, developmental delay, Down Syndrome, global developmental delay	Intellectual disability, autism, developmental delay, learning	Intellectual/learning
Physical	Spinal cord injury, cerebral palsy, other physical	Physical	Amputation, skin disorders & burns, endocrine & immune system, urogenital disorder, reproductive system, gastro-intestinal system, circulatory system, respiratory system, cancer, chronic pain, muscular, skeletal & connective tissue
Psychosocial	Psychosocial disability	Psychiatric	Psychological/psychiatric, nervous system
ABI	Acquired brain injury, stroke	Acquired brain injury	Acquired brain injury
Other	Multiple sclerosis, other neurological	Neurological	Congenital abnormalities, poorly defined cause, infectious disease, inherited disorders, granted prior to 12/11/91

Appendix 4. Technical specifications of the disability indicators

This technical appendix describes the methodology used to derive the disability indicators in each dataset, including the files used to derive the disability indicators, the variables used (including coding), and any aggregation of data. ‘Disability items’ are those used for the analytic cohort in this report. ‘Additional disability items’ are included in the dataset linked with MADIP for further testing.

National Disability Insurance Scheme (NDIS)

Two NDIS data files were used to identify disability:

- *ndda_or_accessrequests_2006_vrf* to identify start and end date of NDIS plans (‘access requests file’)
- *ndda_or_participantdemo_2006_vrf* for demographics and disability information (‘participant demographics file’)

The NDIS dataset contained records from 1 July 2013 to 30 June 2020. The dataset contained 586,581 individuals, including 362,698 individuals who were NDIS participants (i.e., had an NDIS plan) at any time during that period.

From the participant demographics file, information was extracted about date of birth, date of death, age, sex, Indigenous status, cultural and linguistic diversity, country of birth, language spoken at home, state of residence, sa2 and remoteness and disability (see Table A1). The disability information was used to generate indicators of (1) overall disability, (2) severe disability, (3) disability groups (combining data from primary and secondary disability groups). These data were then merged with the access requests file to identify start and end dates of plans for each individual. Note, people were only classified as having a disability if they were NDIS participants, defined as having a plan start date.

Table A1. NDIS variables

	Coding	Variables used
AIHW_PPN	Unique ID	AIHW_PPN
Demographics		
dob	01monthYYYY	BrthDt_MMYYYY
dod	01monthYYYY	DeathDt_MMYYYY
age_2018	derived from month and year of birth to derive a variable specifying age at end of financial year 2017/2018	BrthDt_MMYYYY
sex	1=male, 2=female, .m=unspecified	GndrTyp
indigenous	Indigenous status: 1=First Nations Autralians, 0=not indigenous, .m=missing	ATSISts
cald	1=CALD, 0=Not CALD, .m=not stated	CALDSts
cob	String variable with list of countries	CntryNm
language	String variable with list of languages	LangNm
state	State of residence: 1=NSW, 2=VIC, 3=QLD, 4=SA, 5=WA 6=TAS 7=NT 8=ACT 9=Other territory .m=missing	RsdsInStateCd
sa2	Numeric	SA2Cd2016
remoteness	1=major city, 2=>50k, 3=15k-50k, 4=5k-15k, 5=<5k, 6=remote, 7=very remote, .m=missing	RsdsInMMMCd
Disability items		
disability_ndis	1=ndis participant, 0=not ndis participant	FrsrPlanAprvlDt
disability_group	1=S/S (Hearing; Visual; Other s/s)	NDISDsbltyGrpNm,

	2= intellectual/learning (ID; Autism; Down Syndrome; Developmental Delay; Global Developmental Delay) 3= physical (Spinal cord injury; cerebral palsy; other physical) 4=psychosocial (psychosocial disability) 5=ABI (ABI; Stroke) 6=Other including neurological (Other; MS; Other neurological) .m=missing	
severe	1=severe or profound, 0=other, .m=missing	SVRTYSCR (severe defined as SVRTYSCR=11-15)
start_date	DDmonthYYYY	FrsrPlanAprvlDt
end_date	DDmonthYYYY	Exit_month
secondary_disability	1=yes, 0=no	OtherDsblty,
secondary_disability_ss	1=S/S (See categorisation above), 0=not S&S	OtherDsblty
secondary_disability_int	1=intellectual/learning, 0=not intellectual/learning	OtherDsblty
secondary_disability_phys	1=physical, 0=not physical	OtherDsblty
secondary_disability_psych	1=psychosocial, 0=not psychosocial	OtherDsblty
secondary_disability_abi	1=ABI, 0=not ABI	OtherDsblty
secondary_disability_other	1=other, 0=not other	OtherDsblty
Additional disability items		
request_status	Access met; access not met; cancelled; in progress; access revoked/ceased; withdrawn	AcsRqstDcsnAdj
request_date	DDmonthYYYY	AcsRqstDcsnDt
request_reason	Disability met; disability not met;	AcsRqstDcsnRsn
exit_reason	Deceased; access revoked; disability/early intervention not met; moved to community/aged care and 65+; participant requested; residence not met	ExitRsnDesc

Disability Services National Minimum Dataset (DS NMDS)

Two of the data files were used in our analysis:

- *outcomes_sr_sto* to identify start and end date of disability services (‘services received file’)
- *outcomes_su* for demographics and disability information (‘service users file’)

The DS NMDS dataset contained records from 1 July 2010 to 1 June 2019. The dataset contained 672,654 individuals, all of whom used disability services and therefore contributed to the disability cohort.

From the ‘service users file’, information was extracted about age, sex, Indigenous status, state of residence, and disability (see Table A2). The disability information was used to generate indicators of (1) overall disability, (2) severe disability, (3) disability groups (combining data from primary and secondary disability groups). These data were then merged with the ‘services received file’ to identify start and end dates of services for each individual.

Within each financial year, individuals could have multiple records with different start and end dates, relating to the use of different service types, provided by different agencies, and in different States. We consolidated the records within each financial year using the earliest start date and the latest end date within the financial year to describe the period during which the individual was receiving services.

Table A2. DS NMDS variables

	Coding	Variables used
AIHW_PPN	Unique ID	AIHW_PPN
year	2011-2019 (relating to financial years 2010/2011 to 2018/2019)	year
Demographics		
sex	1=male, 2=female, .m=missing	su_sex
age_jun30	Numeric (note, a few negative values)	su_age_jun30
age_2018	derived from age at end of each financial year to derive a variable specifying age at end of financial year 2017/2018	su_age_jun30
indigenous	1=First Nations Australians, 0=not Indigenous, .m=missing	su_indigenous
cob	String variable	su_country_birth
interpreter	1=yes – language, 2=yes – non-spoken communication, 3=no, .9=missing	interpreter_required
state	State of residence: 1=NSW, 2=VIC, 3=QLD, 4=SA, 5=WA 6=TAS 7=NT 8=ACT 9=Other territory .m=missing	su_asgs_from_postcode
sa2	Numeric	su_SA2_maincode_2016
Disability items		
disability_dsnmds	1=used disability services (whole dataset coded as 1)	
disability_group	1=S/S (deafblind; vision; hearing; speech) 2=intellectual/learning (intellectual; learning; autism; developmental delay) 3=physical (physical) 4=psychosocial (psychiatric) 5=ABI (ABI) 6=Other including neurological (neurological) .m=missing	su_primary_disability
severe	1=severe or profound, 0=other, .m=missing	su_support_selfcare su_support_mobility su_support_comm (if any coded 1/2)
start_date	DDmonthYYYY	sr_service_start_mm_yyyy
end_date	DDmonthYYYY (derived from two variables, using the earlier date when they differed, as recommended by DS NMDS contact)	sr_service_exit_mm_yyyy sr_last_service_date
secondary_disability	1=yes, 0=no	From secondary disability group derived variables
secondary_disability_ss	1=S&S (See categorisation above), 0=not S&S	su_osd_deafblind su_osd_vision su_osd_hearing su_osd_speech
secondary_disability_int	1=intellectual/learning, 0=not intellectual/learning	su_osd_intellectual su_osd_specific_learn su_osd_autism su_osd_dev_delay

secondary_disability_phys	1=physical, 0=not physical	su_osd_physical
secondary_disability_psych	1=psychosocial, 0=not psychosocial	su_osd_psychosocial
secondary_disability_abi	1=ABI, 0=not ABI	su_osd_abi
secondary_disability_other	1=other, 0=not other	su_osd_neurological

Data Over Multiple Individual Occurrences (DOMINO)

The DOMINO dataset contained records from July 2009 to June 2019. The dataset contained 5,701,477 individuals, including 2,922,599 individuals who contributed to the disability cohort.

Four types of data files were used to identify disability in our analysis:

- payment, allowance and benefit files : *det_ben, pyh_reg, pyh_oneoff* and *pyh_third* to identify disability related payment allowances and benefits in a given fiscal year.
- medical files*: mcd_dtls* to identify medical conditions of DSP recipients and *mcd_caree* to identify medical condition of those receiving care from those receiving carer payment or carer allowance.
- education files: *edu_ftb* and *edu_dtls* to identify individuals studying in particular types of institutions.
- Healthcare card file: *det_hcc* to identify carees (children of the carers).

Other files were used to identify demographic characteristics in particular age, gender and Indigenous status (from the file *stt_pit*) as well as state of residence (file *loc_dtls*).

In total, 11 files were used to identify disability and key characteristics of DSS recipients (variables listed in Table A3).

The demographic file contained 5,567,119 individuals, including 2,797,693 individuals identified as having a disability, therefore there was a small proportion of the disability cohort for whom we did not have demographic data.

From the payment, allowance and benefit files, information was extracted about payment types, start and end dates. Individuals receiving either the Disability Support Pension (for at least one day) or Sickness allowances (for at least 180 days) were extracted from the benefit file. From the payment file, we included those receiving payment related to the Disability Wage Supplement, Business Services Wage Assessment Tool, Mobility Allowance or components such as the Youth Disability Supplement (whether these are one-off payments or not). For consistency, only those whose benefits were current were retained in the dataset.

In the medical file *mcd_dtls*, we used variables that identify types of medical conditions. Individuals with a primary condition, a secondary condition, a manifest disability or a period of incapacity were retained in the dataset.

While primary conditions are generally permanent, secondary conditions can be permanent, recurrent or temporary. In the latter case, only conditions that lasted at least six months were kept. The medical condition information was used to generate disability group (see table A3). An individual could have several secondary disability groups if they had several conditions. The disability group was based on the condition which is permanent or has been fully diagnosed, treated and stabilised.

In the medical caree database *mcd_caree*, all individuals were retained under the assumption that if someone receives the carer payment or carer allowance, the caree must have a disability. Carees' medical conditions are classified differently depending on whether the caree is an adult or a child. For a carer to receive a payment for an adult caree, the caree must require constant care and, for this reason, we identified adult carees as having severe disability (s_caree). The information on the types of disability was used to fill the secondary disability groups (but there was no information to enable determination of primary disability group). For children, three variables were present: (1) 'recognised disabilities', (2) 'non-recognised disabilities', both of which indicate serious conditions, and a third variable which picks one of the conditions identified in either of the recognised or non-recognised disability groups, which we used to determine the primary disability group.

Besides medical condition, the general medical file *mcd_dtls* also includes information on manifest disability and temporary incapacity to meet mutual obligations.

A manifest disability is a medical condition that qualifies someone for the receipt of the Disability Support Pension without going through a medical assessment. The list of categories include, among others, permanent blindness and intellectual disability with an IQ less than 70.

We defined temporary incapacity as an episode where someone has either an ongoing partial capacity to work or a reduced capacity to work for a period of time. Ongoing partial capacity to work is recorded in a job capacity assessment or an employment services assessment. Reduced capacity to work is also recorded in a job capacity assessment or an employment services assessment, but the assessment can also be made by a Centrelink customer service officer based on medical evidence (such as a medical certificate) alone. Episodes of temporary incapacity that lasted less than six months were excluded from the dataset.

Lastly we selected individuals attending Special Schools, as identified in the education files, and children that were being cared for, linked to carers with healthcare cards, identified in the healthcare card file.

All medical and non-medical information was used to generate indicators of (1) overall disability in a financial year, (2) having had a disability ever and (3) disability group (combining data from primary and secondary disability groups). All data were merged together and start and end dates for payments, education, and conditions were aggregated by financial year.

Table A3. DOMINO variables

	Coding	Variables used
AIHW_PPN	Unique ID	AIHW_PPN
year	2010-2020 (relating to financial years 2009/2010 to 2019/2020)	year
Demographics		
dob_domino	Date of birth 01MMMYYYY	DATE_OF_BIRTH
dod_domino	Date of death 01MMMYYYY	DATE_OF_DEATH
sex_domino	1=male, 2=female	GENDER
indigenous_domino	1=First Nations Australians, 0=not Indigenous, missing if not asked or not willing to answer	INDIG_CODE
cob_domino	Country of birth	BIRTH_CTRY_CODE
flag_dob	1 if several date of birth identified for the same individual, 0 otherwise	DATE_OF_BIRTH
flag_dod	1 if several date of death identified for the same individual, 0 otherwise	DATE_OF_DEATH
flag_gender	1 if several gender identified for the same individual, 0 otherwise	GENDER
flag_indigenous	1 if several indigenous status identified for the same individual, 0 otherwise	INDIG_CODE
state_domino	Latest state recorded for the individual 1=NSW, 2=VIC, 3=QLD, 4=SA, 5=WA, 6=TAS, 7=NT, 8=ACT	STATE
sa2_domino_static	Last SA2 of residence identified (ASGS 2011)	SA2_MAINCODE_2011
remoteness_domino	1=remote 0=not remote	RMT_IND
stateyr_domino	State where resided longest during financial year. 1=NSW, 2=VIC, 3=QLD, 4=SA, 5=WA, 6=TAS, 7=NT, 8=ACT	STATE
sa2_domino_dynamic	SA2 where resided longest during financial year	SA2_MAINCODE_2011
dob_domino	Date of birth 01MMMYYYY	DATE_OF_BIRTH

Disability items

disability_domino	1 if identified as having a disability in a given financial year, i.e., whether has a medical condition (primary or secondary and if secondary and temporary must last at least 6 months) , has a manifest condition, has an incapacity of at least 1 day, has a terminal illness, received a disability related payment, is identified as a caree, or is going to a special education institution.	disability_group, secdis_any, s_manifest_bli, s_manifest_int, s_manifest_other, s_tempcap, s_jobcap, dis_domino_term, dis_domino_dsp, dis_domino_ska, dis_domino_bswat, dis_domino_mob, dis_domino_yds, dis_domino_careehcc, dis_domino_caree, dis_domino_edu
disability_ever_domino	1 if identified as having a disability in any financial year.	disability
disability_group	1=S&S (deafblind; vision; hearing; speech) 2=intellectual (intellectual; learning; autism; developmental delay) 3=physical (physical) 4=psychosocial (psychiatric) 5= acquired brain injury (ABI) 6=Other including neurological (neurological) [mapping from Domino to these 6 categories in Appendix 3]	MED_PRMY CHLD_MED_CODE
dis_first	1st of July of financial year	PERIOD_START_DATE PERIOD_END_DATE
dis_last	30th of June of financial year	PERIOD_START_DATE PERIOD_END_DATE
secdis_any	1=yes, 0=no	MED_SCNDRY CHLD_RDL_CODES CHLD_NON_RDL_CODES ADLT_MED_CODES
secdis_ss	1=S&S (See categorisation for disability_group), 0=not S&S	MED_SCNDRY CHLD_RDL_CODES CHLD_NON_RDL_CODES ADLT_MED_CODES
secdis_int	1=intellectual, 0=not intellectual	MED_SCNDRY CHLD_RDL_CODES CHLD_NON_RDL_CODES ADLT_MED_CODES
secdis_phys	1=physical, 0=not physical	MED_SCNDRY CHLD_RDL_CODES CHLD_NON_RDL_CODES ADLT_MED_CODES
secdis_psych	1=psychosocial, 0=not psychosocial	MED_SCNDRY CHLD_RDL_CODES CHLD_NON_RDL_CODES ADLT_MED_CODES

secdis_abi	1=ABI, 0=not ABI	MED_SCNDRY CHLD_RDL_CODES CHLD_NON_RDL_CODES ADLT_MED_CODES
secdis_other	1=other, 0=not other	MED_SCNDRY CHLD_RDL_CODES CHLD_NON_RDL_CODES ADLT_MED_CODES
Additional disability items		
s_caree	1 if identified as adult caree in the caree medical file. 0 otherwise	Identifier AIHW_PPN
s_manifest_bli	1 if identified as permanently blind, one of the manifest medical conditions for eligibility to DSP. 0 otherwise	MAN_CODE
s_manifest_int	1 if identified as having a manifest intellectual disability with IQ<70, one of the manifest medical conditions for eligibility to DSP. 0 otherwise	MAN_CODE
s_manifest_other	1 if Identified as having a manifest medical conditions for eligibility to DSP which is not permanent blindness and not an intellectual disability with IQ<70. 0 otherwise	MAN_CODE
s_tempcap	number of days of temporary/ reduced capacity episodes during financial year, conditional on the episode lasting at least 6 months across financial years.	INCAP_START, INCAP_END
s_tempcaphrs	lowest weekly hours during long term incapacity/temporary reduced capacity episodes occurring during the financial year	INCAP_WK
s_jobcap	number of days of partial capacity to work episodes during financial year	CURR_CAPCTY, WITH_INT
s_jobcaphrs	Lowest weekly hours during partial capacity to work episodes in the financial year	CURR_CAPCTY, WITH_INT
dis_domino_perm	1 secondary condition is permanent (and/or fully treated), recurring or temporary (only if 6months plus), 0 otherwise	MED_SCNDRY_PERM
dis_domino_term	1having a terminal illness in medical caree file (DOMINO), 0 otherwise	TERM_ILL
dis_domino_dsp	1 if received Disability Support Pension Payment during the financial year, 0 otherwise	BEN_TYPE_CODE
dis_domino_ska	1 if received Sickness allowance payment for at least 180 days and one of those days occur during the financial year, 0 otherwise	BEN_TYPE_CODE
dis_domino_bswat	1 if received the business services wage assessment tool during the financial year, 0 otherwise	CMPNT_TYPE
dis_domino_mob	1 if received the mobility allowance during the financial year, 0 otherwise	BEN_TYPE
dis_domino_yds	1 if received the Youth Disability Supplement during the financial year, 0 otherwise	CMPNT_TYPE
dis_domino_careehcc	1 if identified as a caree in the concession card dataset during the financial year, 0 otherwise	CDA CHILD

dis_domino_caree	1 if identified as a caree in the carer payment dataset during the financial year, 0 otherwise	Identifier AIHW_PPN
dis_domino_edu	1 if going to a special institution, 0 otherwise	INST_TYPE

Medicare Benefits Schedule (MBS)

One file was used to identify people with disability:

- *eo2020_3_1185_mbs_2010_2020* was used to identify disability and start and end dates ('MBS file')

The MBS file contained records from 1 January 2010 to 30 June 2020. The dataset contained 4,759,215 individuals, including 127,377 individuals identified as having disability based on MBS items.

From the MBS file, information was extracted about broad type of Medicare service and date of service to generate indicators of: (1) overall disability, and (2) disability group (intellectual disability, Table A4).

Codes relating to provision of autism, pervasive developmental disorder or disability services by allied health professionals (MBS group: M10; MBS codes: 82000-82035), early intervention services for children with autism, pervasive developmental disorder or disability (MBS group: A29; codes: 135-139) and consultant psychiatrists attendance for assessment and diagnosis of autism or any other pervasive development disorder for children aged under 13 years and development of a treatment and management plan (MBS group: A08; code: 289) were used to identify overall disability. These were associated with dates of service. Only services in the A08 MBS group (code 289) were used to identify disability group (intellectual or learning disability) as all other services were available to children with a range of different disabilities. We did not include codes relating to health assessments (MBS groups: AN.0.36 and AN.7.5, codes: 224-227 and 701-707) because these items are not specific to people with disability. No information was available on demographics, state of residence, severity or secondary disability.

Records were aggregated by financial year because some individuals in the dataset had multiple records for these specific codes. We consolidated these records into a single record per financial year, using the first record in that financial year as a start date and the last record in the financial year as an end date. We created a count of number of services by treatment group for each individual by financial year.

Table A4. MBS variables

	Coding	Variables used
AIHW_PPN	Unique ID	AIHW_PPN
year	Numeric, 2010-2020 (relating to financial years 2009/2010 to 2019/2020)	Date_of_service
Disability items		
disability_mbs	1=disability, 0=no disability	Current_item
Disability_group_mbs	2=intellectual/learning disability	Current_item
start_date	Date of first service within the financial year: DDmonthYYYY	Date_of_service
end_date	Date of last service within the financial year: DDmonthYYYY	Date_of_service
Additional disability items		
treat_group1_count	Count of M10 services by financial year	Current_item
treat_group2_count	Count of A29 services by financial year	Current_item
treat_group3_count	Count of A08 services by financial year	Current_item

Specialist Homelessness Services Collection (SHSC)

Two data files were used to identify people with disability:

- *shsc_client* for demographics ('client file')
- *shsc_sp* to identify disability, severity and start and end dates ('support period file')

The SHSC dataset contained records from 1 July 2010 to 1 June 2020. The dataset contained 871,065 individuals, including 78,162 individuals with a disability.

From the client file, information was extracted about Indigenous status (see Table A5). Although the client file contained a disability flag, this variable was not used because there was more detail about disability relating to different time periods in the support period file. The support period file contained records relating to episodes of support, and for each episode information was collected on core activity limitations. We used this information to derive an indicator of (1) overall disability and (2) severity, using start date of services to identify when disability data were collected. No information was collected on disability group or secondary disability.

Some individuals in the dataset had multiple records, describing multiple episodes using homelessness and emergency accommodation services. We consolidated these records into a single record per financial year, using the latest record in that financial year as a start date to identify the most recent disability data. Information on disability and severity was recorded for each financial year. Where disability information differed between different records in a single financial year, the most recent disability information was applied for the financial year. In the case of multiple records with the same start date, the most severe disability information was used if it different between records.

Table A5. SHSC variables

	Coding	Variables used
AIHW_PPN	Unique ID	AIHW_PPN
year	Numeric, 2011-2020 (relating to financial years 2010/2011 to 2019/2020)	Date_of_service
Demographics		
Indigenous	1=First Nations Australians, 0=not Indigenous, .m=missing	indigenous_er
Disability items		
disability_shsc	1= disability, 0=no disability, .m=missing	dis_selfcare, dis_mobility, dis_comm (if any coded 1/3)
severe	1=severe or profound, 0=other, .m=missing	dis_selfcare, dis_mobility, dis_comm (any coded 1)
start_date	Of homelessness services: DDmonthYYYY, .m=missing	episode_start_month, episode_start_year
Additional disability items		
need_disab_services	1=yes, 0=no	need7
provision_disab_services	1=yes, 0=no	prov7
referred_disab_services	1=yes, 0=no	ref_7

Public Housing and State Owned and Managed Indigenous Housing (PH and SOMIH)

One data file was used to identify people with disability:

- *eo2020_3_1185_ph_somih_20210511* for demographics, disability and start date ('ph file')

The PH and SOMIH dataset contained records from the financial years 2017/2018, 2018/2019 and 2019/2020. The dataset contained 551,197 individuals, including 113,217 individuals with a disability.

Information was extracted about Indigenous status, disability status and the start date of public housing or Indigenous housing (see Table A6). No information was collected on disability group, severity or secondary disability. If individuals in the dataset had more than one record per financial year, the most recent record was used to use the most recent disability data.

Table A6. PH and SOMIH variables

	Coding	Variables used
AIHW_PPN	Unique ID	AIHW_PPN
year	Numeric, 2018-2020 (relating to financial years 2017/2018 to 2019/2020)	Date_of_service
Demographics		
indigenous	1=First Nations Australians, 0=not Indigenous, .m=missing	indig_p
state	1=NSW, 2=VIC, 3=QLD, 4=SA	state
sa2	numeric	sa2_maincode_2016
Disability items		
disability_ph	1= disability, 0=no disability, .m=missing	dis_p
start_date	Of tenancy: DDmonthYYYY	ten_start_month, ten_start_year
Additional disability items		
hh_disab	1= disability, 0=no disability	dis_h

Pharmaceutical Benefits Scheme (PBS)

PBS items for medications that may indicate health conditions associated with disability (N05A, N05B, N05C, NO6A and N06B) are potentially useful for distinguishing subgroups of people with disability. The file was too large to be included as a single file so was separated into two distinct files:

- *Outcomes_pbs_2010_2015* including data on prescription of the listed PBS items from years 2010 to 2015
- *Outcomes_pbs_2016_2020* including data on prescription of the listed PBS items from years 2016 to 2020

The PBS dataset contained records from 1 January 2010 to 30 June 2020. The dataset contained 3,110,909 individuals who received prescriptions for the listed PBS items, including 597,799 individuals with a disability who were identified in other datasets.

From the PBS file, information was extracted about prescription of certain types of nervous system drugs to generate an indicator of disability group (psychosocial disability, Table A7).

Codes relating to prescription of antipsychotics (codes N05AA01 to N05AX13) were used to identify disability group (psychosocial disability). Codes relating to the prescription of anxiolytics (codes N05BA01 to N05BE01), hypnotics and sedatives (codes N05CD02 to N05CF01), antidepressants (codes N06AA02 to N06AX23) and psychostimulants (codes: N06BA02 to N06BA12) were retained as additional disability items to test their capacity to identify subgroups of people who may have psychosocial disability. Codes were associated with dates of supply, which were used to identify start dates. No information was available on demographics, state of residence, severity or secondary disability.

Records were aggregated by financial year because some individuals in the dataset had multiple records relating to these codes. We consolidated these records into a single record per financial year, creating a start and end date for each PBS drug type using the first prescription within the financial year as a start date and the last prescription as an end date. We created a count of number of prescriptions by PBS drug type for each individual by financial year.

Table A7. PBS variables

	Coding	Variables used
AIHW_PPN	Unique ID	AIHW_PPN
year	Numeric, 2010-2020 (relating to financial years 2009/2010 to 2019/2020)	Date_of_Supply
Disability items		
disability_group_pbs	4=psychosocial	ATC (codes N05AA01 to N04AX13)
Additional disability items		
drug_type1_start_date	Date of first N05A prescription within the financial year: DDmonthYYYY	Date_of_Supply
drug_type1_end_date	Date of last N05A prescription within the financial year: DDmonthYYYY	Date_of_Supply
drug_type1_count	Count of N05A prescriptions by financial year	ATC
drug_type2_start_date	Date of first N05B prescription within the financial year: DDmonthYYYY	Date_of_Supply
drug_type2_end_date	Date of last N05B prescription within the financial year: DDmonthYYYY	Date_of_Supply
drug_type2_count	Count of N05B prescriptions by financial year	ATC
drug_type3_start_date	Date of first N05C prescription within the financial year: DDmonthYYYY	Date_of_Supply
drug_type3_end_date	Date of last N05C prescription within the financial year: DDmonthYYYY	Date_of_Supply
drug_type3_count	Count of N05C prescriptions by financial year	ATC
drug_type4_start_date	Date of first N06A prescription within the financial year: DDmonthYYYY	Date_of_Supply
drug_type4_end_date	Date of last N06A prescription within the financial year: DDmonthYYYY	Date_of_Supply
drug_type4_count	Count of N06A prescriptions by financial year	ATC
drug_type5_start_date	Date of first N06B prescription within the financial year: DDmonthYYYY	Date_of_Supply
drug_type5_end_date	Date of last N06B prescription within the financial year: DDmonthYYYY	Date_of_Supply
drug_type5_count	Count of N06B prescriptions by financial year	ATC