Interim learnings from test case analyses

September 2021

National
Disability
Data Asset

EXECUTIVE SUMMARY

This second test case report from the National Disability Data Asset (NDDA) Pilot presents information on methodologies used to identify disability indicators and highlights key data gaps in administrative data.

This report aims to provide NDDA governance forums, the disability community and senior stakeholders with information about how effective the approach used during the Pilot was in detecting people with disability compared to the Survey of Disability, Ageing and Carers (SDAC) that is currently relied on but occurs every three years. It also highlights some data development work undertaken during the pilot phase; work required to continue to improve data quality and shares some insights from the individual test cases. A third report that compiles more comprehensive analytical insights from each test case will be finalised at the end of 2021.

Key findings:

- The NDDA pilot has laid foundations for development of a suite of disability indicators that will form the cornerstone of an enduring disability data asset.
    - Within the test cases, the indicators developed have significantly enhanced the ability to "identify" people with disability in the data within, and across, service systems.
    - Disability indicators based on the linked data aligned closely with Survey of Disability, Ageing and Carers (SDAC) estimates for adults aged 25-64.
- Disability for children and young people can be more reliably ascertained through use of health and education records
- Under-reporting of disability is particularly prevalent for First Nations Australians
    - First Nations people are over-represented in both disability and justice system populations. Research led by First Nations Australians, and investments in data development, are needed to address this gap and design appropriate frameworks for conceptualising disability, supporting the needs and outcomes for First Nations Australians in the enduring asset.
- Further work is required on:
    - Development of a reliable indicator of disability for older people and for children and young people
    - Data development around disability related to complex trauma
    - Ongoing data quality improvements for National Minimum Datasets
    - Development of comprehensive metadata
    - Ensuring data is fit-for-purpose
- National Minimum Data Sets (NMDS) provide a valuable source of nationally collated state data but were found to have limitations in some cases. Data sourced directly from states were generally found to provide richer and more granular sources of information.

However, benefits need to be balanced against the challenges of collating nationally comparable and standardised data under this approach as part of the enduring asset.

- Ongoing investments in data improvement are required to support the enduring NDDA. In the short term, better quality metadata (and data dictionaries) can facilitate the production of quality insights.
    - Streamlining data improvement efforts across different initiatives will have benefits far beyond individual projects. For example, conversations are underway with DSS to share data quality learnings from the NDDA pilot and ensure alignment with improvement efforts related to Australia's Disability Strategy. If agreed and funded, data remediation work undertaken as part of the NDDA could also likely inform other priority national initiatives, such as Closing the Gap.
    - Data remediation work undertaken as part of the NDDA pilot has enabled exploration of victimisation rates among people with disability for the first time. These insights were identified as a critical gap by the Royal Commission into Violence, Abuse, Neglect and Exploitation of People with Disability, and demonstrates the value of investment in data improvement to specific areas of policy, safeguarding and service delivery.
- The technical design for the enduring asset will need to have the capability to manage very large volumes of data and computationally intensive analysis, to enable refinement and derivation of high-quality, reliable disability indicators.
- The group of people identified by disability indicators derived from linked administrative data will be affected by:
    - the use of different definitions of disability in different administrative data sources, which align to a greater or lesser extent with key population surveys such as SDAC;
    - the collection of different types and depths of information on people's health conditions, impairments and functioning, including self-reported and other means;
    - the coverage of different datasets, including factors determining eligibility for services and supports (which may vary over time with changes in policy); and
    - the varying frequency of collecting or reporting disability data, rationale for collection and method of collection (including electronic capture or manual data entry).
- Careful consideration is needed of the validity and utility of a suite of derived disability indicators being applied for specific purposes.

Contents

Glossary

Acronym	Description
AAIP	Australian Apprenticeships Incentives Program
ADS	Australia's Disability Strategy
AEDC	Australian Early Development Census; data is collected every 3 years
APC	NSW Annual (Community) Preschool Census
BOCSAR	NSW Bureau of Crime Statistics and Research
BSWAT	Business Wage Assessment Tool; this payment scheme closed in December 2018
Census	Australian Census of Population and Housing; 2016 is the most recent enumeration period for this dataset
CMI-ODS	Public clinical mental health dataset
DIP-HLSN	NSW Disability and Inclusion Program – Higher Learning Support Needs
DOMINO	Data Over Multiple INdividual Occurrences *DOMINO is a dataset held by the Department of Social Services which contains payments made to individuals*
DRC	Royal Commission into Violence, Abuse, Neglect and Exploitation of People with Disability
DS-NMDS	Disability Services National Minimum Dataset *Recipients of specialist disability services provided under historic National Disability Agreements*
DSP	Disability Support Pension
ESD-IF	NSW DoE Student Disability Data collection: Integrated Funding Support
ESD-SC	NSW DoE Student Disability Data collection: Support Class
HEIMS	Higher Education Information Management System (South Australia)
ICF	WHO International Classification of Functioning, Disability and Health
LGBTIQ+	an evolving acronym that stands for lesbian, gay, bisexual, transgender, intersex, queer/questioning, asexual. Many other terms (such as non-binary and pansexual) that people use to describe their experiences of their gender, sexuality and physiological sex characteristics.
MBS	Medicare Benefits Schedule
MSMA	Multi-Stage Median Algorithm approach *described in the "National Best Practice Guidelines for data linkage activities relating to Aboriginal and Torres Strait Islander People" (ABS & AIHW, 2012)*
MYC-EI	NSW Mid-Year Census: Early Intervention
MYC-GP	NSW Mid-Year Census: Government Preschools
NCCD	Nationally Consistent Collection Data on School Students with Disability
NDDA	National Disability Data Asset
NDIS	National Disability Insurance Scheme;
NMDS	National Minimum Data Set
PBS	Pharmaceutical Benefits Scheme
PDSP	NSW Preschool Disability Support Program
PH & SOMIH	Public Housing and State-Owned and Managed Indigenous Housing

Acronym	Description
SDAC	Survey of Disability, Ageing and Carers *In 2018 there were 65,805 survey responses collected. This resulted in an estimate of 4.4 million people living with disability in Australia (17.7% of the population). SDAC excludes people living in very remote areas or discrete Aboriginal or Torres Strait Islander communities. SDAC provides data at the national level and at the state level for New South Wales, Victoria, Queensland and Western Australia. Data for other states and territories may be limited due to smaller sample size.*
SHSC	Specialist Homelessness Services Collection
TVA	Total Vet Activity
VEMD	Victorian Emergency Minimum Dataset *Presentations at Victorian public hospitals with designated emergency departments*
VAED	Victorian Admitted Episodes Dataset *Admitted patient episodes from Victorian public and private hospitals*

1 Background

The NDDA Pilot included five test cases:

1. Outcomes measurement in housing
2. Services and supports for people with disability and mental health issues
3. Pathways from education to employment
4. Experiences with the justice system
5. Early childhood supports

Some test cases are significantly more progressed (or completed) in their analytics work at the time of this report. The third and final Pilot report due in November 2021 will focus on comprehensive presentation of results of policy analysis across all test cases.

2 Conceptualisation and measurement of disability

The current best available information source for measuring disability prevalence in Australia is considered to be the Australian Bureau of Statistics' Survey of Disability, Ageing and Carers (no. 4430.0) (SDAC)[1], which aligns closely with the WHO International Classification of Functioning, Disability and Health (ICF). The ICF is the international standard framework and classification for organising and documenting information about functioning and disability and conceptualises functioning as a "dynamic interaction between a person's health condition(s), environmental and personal factors". The SDAC measurement approach aligns with the ICF model, with the resulting disability cohort comprising people who experience activity limitations or participation restrictions associated with health conditions. While SDAC has strong alignment with the ICF framework, there are inherent limitations involved in generalising outputs arising from applying a survey methodology, notably challenges in generating insights at local geographic levels and for intersectional groups such as LGBTIQ+ and culturally and linguistically diverse communities. Data linkage using administrative data provides the opportunity to add value in these areas.

Key statistics for disability prevalence in Australia (Survey of Disability, Ageing and Carers, 2018)

- In 2018 there were 4.4 million Australians with disability (17.7% of the population).
- The prevalence of disability increased with age – one in nine (11.6%) people aged 0-64 years and one in two (49.6%) people aged 65 years and over had disability.

---

[1] SDAC is a survey methodology; its 2018 estimate of 4.4 million people with disability in Australia is based on a sample of 65,805 respondents. SDAC excludes people living in very remote areas or discrete Aboriginal or Torres Strait Islander communities. SDAC provides data at the national level and at the state level for New South Wales, Victoria, Queensland and Western Australia. Data for other states and territories may be limited due to smaller sample size.

- 5.7% of all Australians had a profound or severe disability, defined as sometimes or always needing help with daily self-care, mobility or communication activities.
- 11.2% of all Australians had a physical disability, 6.2% had a sensory or speech disability, and 4.6% had a psychosocial disability.

Disability prevalence among Aboriginal and Torres Strait Islanders (National Aboriginal and Torres Strait Islander Health Survey, 2018)

- Almost four in ten (37.6%) Aboriginal and Torres Strait Islanders had disability*.
- 8% of Aboriginal and Torres Strait Islanders had profound or severe disability.

**Note: the National Aboriginal and Torres Strait Islander Health Survey (NATSIHS) uses a different module to capture disability than SDAC, which tends to produce higher estimates of disability.*

While administrative data offer significant advantages over survey-based methodologies in terms of coverage, existing indicators of disability in administrative data are often inconsistent across data sources and between jurisdictions, posing challenges in "identifying" people with disability in mainstream data. For example, some of the information on disability will be self-reported. Other information will be recorded according to a range of pre-determined, usually policy or program related, criteria which apply to an individual at some point in time (e.g. to assess eligibility for supports or payments). This poses challenges with sourcing data which aligns with the indicators deemed relevant to particular outcomes (for example, against Australia's Disability Strategy). Combining disability information across available sources, when done carefully and in a methodologically sound manner, enables triangulation of findings and increases confidence in results.

Each test case in the NDDA pilot is focused on a unique population of interest and has therefore adopted a different approach to "identify" people with disability. This has provided a good opportunity to assess the suitability of approaches and combinations of data to "identify" disability (including, where relevant, disability type or severity) and allow comparisons of methodologies. All the test cases have assessed the quality of their disability indicators against SDAC population estimates. Furthermore, the Outcomes test case is directly comparing its derived indicators against individual SDAC records; the results of which will be presented in the end-of-year report.

All the test cases included[2] three ‘core’ disability service data sources:

Data source	Description
Data Over Multiple Individual Occurrences (DOMINO)	Recipients of social security payments (e.g., Disability Support Pension)
Disability Services National Minimum Dataset (DS-NMDS)	Recipients of specialist disability services provided under historic National Disability Agreements
National Disability Insurance Scheme (NDIS) data	Participants in the NDIS

The DOMINO data made the largest contribution to the resulting disability indicators, except for children or young people. Each test case supplemented ascertainment with specific domain data relevant to their policy question and population of interest (for example, health or education records); detailed breakdowns of the contributions of source datasets are provided within the test case descriptions (Section 6).

Disability characteristics

The academic literature identifies a wide range of factors around disability which have an impact on outcomes. For instance, some factors such as the types and severity of disability, age of onset, presence of multiple disabilities (i.e. complexity) and chronic medical conditions are often correlated with outcomes. In other instances, social determinants such as poor housing, low income, lower educational attainment, unsafe neighbourhoods, and geographic location are correlated with outcomes. Another important aspect of disability is dynamic: it can resolve or increase over time. Furthermore, what is recorded as the primary disability in the data for an individual may not actually be the one that exerts the most impact on the outcomes of interest. The conditions in which people with disability live, learn and work, household, family and community interactions, and circumstances such as poverty shape the outcomes of people with disability.

The enduring NDDA will need to comprise a suite of high quality, methodologically rigorous indicators fit for purpose to enable appropriate analysis. Ongoing investments in data

---

[2] Although the Early Childhood test case included NDIS and DS-NMDS records in ascertainment of disability, other sources of information (e.g. medical and educational records) were more appropriate for identifying disability among children aged less than 7 years. The Education to Employment test case focused on a year 10 student population and therefore based disability indicators on information captured in school enrolment data (supplemented with information from the three core disability service datasets). The indicator derived for the Justice test case goes beyond individuals who receive disability specific supports (i.e. beyond the core datasets).

improvement are required to support the production of quality insights that can inform policy and improve the lives of people with disability. In addition, commitments will need to be made by participating governments to make changes to their data systems and how they collect data to adopt this good practice.

Disability subtypes

Medical diagnosis codes are captured in DOMINO using the WHO International Classification of Diseases framework (Australian Modification; ICD-10-AM) and provide information on the type of (clinically-diagnosed) disability. For the NDDA pilot, medical diagnosis codes in DOMINO were provided as broad medical groupings only. This may result in some false positives and false negatives in identification of disability subtypes in the test cases, impacting the quality and reliability of disability indicators in some cases.

Given the reliance on DOMINO in the identification of disability, the data provided to the enduring asset should be as granular as possible to ensure the accuracy, reliability and utility of the resulting indicators. This will have impacts on the computing capacity and storage needs of the enduring analytics platform, and it will be critical to design scalable technical infrastructure to support these needs.

Severity

Administrative data generally lack indicators of disability severity. Where present, severity may not be measured consistently across datasets or be comparable with other sources, such as survey data. For example, severity indicators in NDIS records do not align with the SDAC measures of severe or profound disability. Additionally, within a dataset, severity may be captured for some beneficiaries (e.g. recipients of disability support pension payments) but not others, leading to gaps in understanding of severity of disability for significant proportions of the identified disability population. Further work will be needed to harmonise measures of severity across data holdings and develop robust and consistent indicators.

Severity itself is a difficult concept to define and can fluctuate over time, particularly in relation to certain conditions (e.g. mental health). As such, it can be difficult (or inappropriate) to capture in administrative data collections. Linkage to data held outside of government, for example primary health care data, informal care, service provider records, and community support information, may help address this gap.

Temporal associations

The available data has limited capacity to measure the dynamic nature of disability (for example, onset and duration (for non-permanent disability) over time or changes to severity), or ascertain acquired disabilities. Further investment will be required to enhance data collection that enables this measurement.

First Nations Australians

Under-reporting (and under-supporting) of disability is particularly prevalent for First Nations Australians. Administrative data does not currently capture many critical factors for First Nations peoples, such as complex intergenerational trauma and community and family supports. Furthermore, many traditional languages do not have a word for ‘disability’; Aboriginal and Torres Strait Islander communities talk about individuals in enabling, not deficit, terms.

Research led by First Nations Australians is needed to address this gap and design an appropriate framework for conceptualising disability, support needs and outcomes for First Nations people in the enduring asset. Investments in data development are also required across a wide range of data sources, particularly community-controlled organisations where First Nations Australians are more likely to identify and accept a targeted service or support.

The Justice test case (Section 6.4) presents an example of how an Aboriginal Perspectives Expert Panel was embedded to support the analytic work, interpret findings, and identify data development areas particular to First Nations people. This model proved effective not only in guiding the test case, but also informing the data development plan and enduring design.

3 Other emerging insights

All of the test case teams expressed a need for more informative metadata or data dictionaries. In some cases, no documentation was provided with the datasets; without clarity on the context in which data was collected, or the definitions used, this poses a risk to the quality of analysis and interpretation of the results. An enduring asset will only be successful if there is transparency around the underlying data collection, data quality and reliability in the quality and interpretation of the emerging insights.

Many test case teams experienced technical issues with access to both the AIHW and ABS analytic environments, particularly during the onboarding phase. Although most of the issues have been resolved, these led to delays in test case progression. The June test case learnings report previously described the challenges with computing capabilities within the analytic environments, and the need for the test case teams to find workarounds to progress the analysis. The design of the enduring asset should anticipate system demands for future work and ensure the infrastructure is in place to support these needs.

4 Implications

Data insights

National Minimum Data Sets (NMDS) provide a valuable source of nationally collated state and territory data but were found to have limitations in some cases. The quality, consistency and data improvement of these collections will need to be invested in for the enduring NDDA. Data sourced directly from states were generally found to provide richer and more granular sources of information however the potential benefits of this needs to be balanced against the challenges of collating nationally comparable and standardised data under this approach as part of the enduring asset.

Research led by First Nations Australians and investment in data improvement is required to address issues with under-reporting of disability and the conceptualisation of disability relevant to First Nations Australians. It will take time to develop partnerships and frameworks to undertake this work properly. Conversations are underway with the National Indigenous Australians Agency around how to align efforts between the NDDA and other data initiatives.

Analysis work undertaken within the NDDA pilot has laid the foundations for development of a suite of high quality, reliable indicators of disability that will form the cornerstone of an enduring asset.

Data improvement

Ongoing investments in data improvement are required to support the enduring NDDA. Additional details will be provided in the forthcoming Disability Data Development Plan (a NDDA Pilot deliverable). In the short term, better quality metadata (and data dictionaries) can facilitate the production of quality insights.

Streamlining data improvement efforts across different initiatives will have benefits far beyond individual projects. For example, conversations are underway with DSS to share data quality learnings from the NDDA pilot and ensure alignment with improvement efforts related to Australia’s Disability Strategy. If funded, data remediation work undertaken as part of the NDDA could also inform other priority national initiatives such as Closing the Gap.

As part of the NDDA pilot, the Justice test case received funding to develop the first-ever longitudinal dataset on victimisation. In addition to data quality activities, this involved new linkages across individuals, court and offender records. This resulted in high linkage rates (>90%) and for the first time has allowed exploration of victimisation rates among people with disability. Further details are provided in Section 5. As victimisation data was identified as a critical gap by the Royal Commission into Violence, Abuse, Neglect and Exploitation of People with Disability, this demonstrates the indisputable value of investment in data remediation and data linkage.

Analytic capacity

The current analytic environments have struggled with the volumes of data linked for the test cases; in many cases, this has been despite the use of data minimisation techniques (e.g. grouping diagnosis codes by broad categories). More detailed levels of information would enable refinement of the disability indicators. Consequently, to maximise the value of the enduring asset, the technical infrastructure will need to have the capability to manage very large volumes of data and computationally intensive analysis.

In addition to technical capacity, refinement of the disability indicators to incorporate more granular data will require revisions to the methodology developed during the NDDA pilot phase.

5 Data development work undertaken during the pilot phase

As part of this pilot phase the NSW Bureau of Crime Statistics and Research (BOCSAR) was awarded funding to develop a victims cohort dataset. This is the first time that insights on people with disability who were victims of crime reported to police have been produced at this scale. The Royal Commission into Violence, Abuse, Neglect and Exploitation of People with Disability (DRC) has highlighted very limited available data. A DRC-led review of evidence, largely from sample survey data, showed people with disability are more likely to be a victim of a violent crime than those without disability[3]. There is also known to be a strong and complex interaction between victims and offenders – many victims of crimes go on to offend.[4]

Data curation and development allowed NSW Police Force victim records to be linked across unique individuals to provide a longitudinal view of victimisation. The victim data were also linked with court data to determine the outcome of a contact with the justice system for people with disability who are victims, and with offender records to quantify the extent to which people are both victims and offenders. Data linkage rates for the extract were high (>90%), and substantial development activities optimised and assessed the quality of the data. This work identified a cohort of 1.6 million individuals in NSW who were victims. Of these individuals, 19% were also offenders.

---

[3] Centre of Research Excellence in Disability and Health (CRE-DH). Research Report: Nature and extent of violence, abuse, neglect and exploitation against people with disability in Australia. March 2021

[4] Jennings, W. G., Piquero, A. R., & Reingle, J. M. (2012). On the overlap between victimization and offending: A review of the literature. Aggression and Violent Behavior, 17(1), 16–26. doi:10.1016/j.avb.2011.09.003

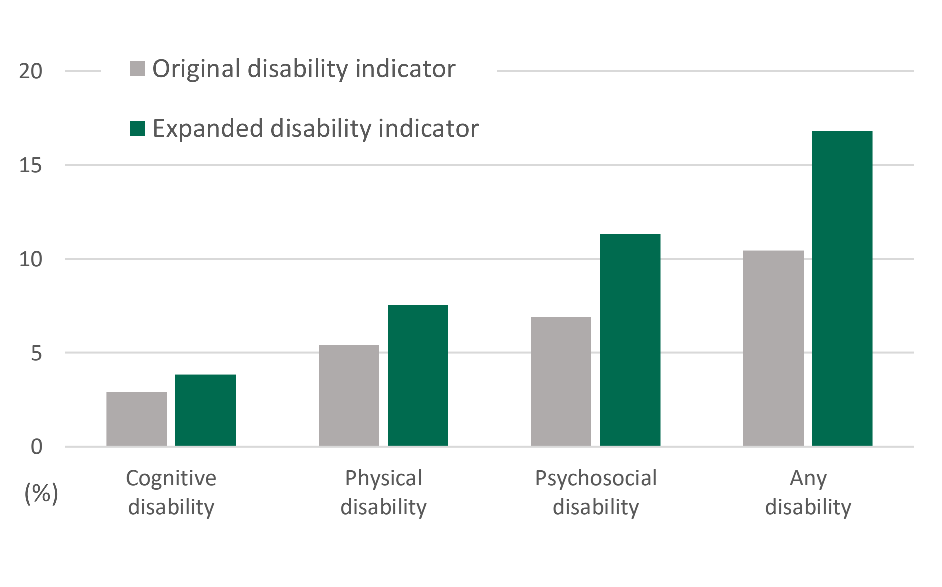
In-scope population for data development work

- Residents of NSW at any time during 2009-2018.
- Aged 10 years or older.
- In contact with criminal justice system as victim.

Two approaches were used to “identify” cohorts of individuals with disability from the linked data. In the first approach (“original indicator”), the disability cohort included people who received a core disability support or service such as NDIS, disability services provided under historic National Disability Agreements, and Disability Support Pension. In the second approach (“expanded indicator”), the cohort included people who may not be in receipt of a core disability support or service but were identified as having disability in the linked data across health, housing and social services based on their diagnoses or services received.

Figure 5.1 shows that the percentage of victims who were identified as having a disability ranged from 10% to 17% depending on whether the original or expanded disability indicator was used.

Figure 5.1. Percentage of victim cohort with disability, by disability type



Nearly 3 in 10 people with disability (according to the original cohort definition) were a victim in contact with the justice system. As this definition only includes reports to the police, it is likely an underestimate of victimisation. From the use of this dataset in this test case some further areas for data development have been identified:

- The identification of First Nations people in this dataset is under-reported (23% missing). Other datasets were used to provide this information, where available. However, concerns remained about whether First Nations people were under-identified in the victim cohort. This will be addressed in the enduring asset by accessing additional data and refining methods to better identify First Nations people.
- The identification of people with disability in the victim dataset may also be an under-representation. Further, the age criteria of disability-specific services and supports (under 65 years) meant that the experiences of people with disability in older age groups could not be examined in this test case. These issues will be addressed in the enduring asset through more uniform flags or methods to identify disability, and with access to wider sources of information, over a longer period of time.

6 Insights from the test cases

6.1 Indicators for service use and outcomes reporting (population-level)

The DSS-led test case *Identification of People with Disability in Linked Administrative Data for Service Use and Outcomes Reporting* (Outcomes Data test case) investigates whether existing administrative data collections can support creation of a comprehensive indicator for people with disability (including those within and outside of the scope of the NDIS and disability support pension [DSP]). It also aims to assess the accuracy of the derived disability indicator and evaluate whether it could be used to reliably report on Australia's Disability Strategy [ADS] / NDIS outcomes for people with disability, as well as services and supports accessed by people with disability. To test the last question, this project looks at housing-related services and supports provided by both disability-specific and mainstream services.

A list of datasets included in this test case is provided in the Appendix.

Highlights and key accomplishments since June 2021

- Held the first workshop of the project's Aboriginal and Torres Strait Islander Perspectives Reference Panel, to seek the panel's perspectives on the identification of people with disability among First Nations Australians and experiences of First Nations Australians with housing services and supports.
- Derived a set of disability indicators based on linked administrative data.
- Commenced analysis of housing-related services and supports, including:
    - Public Housing and State Owned Indigenous Housing;
    - Commonwealth Rent Assistance;
    - Specialist homelessness services; and
    - NDIS-funded specialist disability accommodation services.
- For the first time, analysis across housing supports will use a consistent and more comprehensive disability indicator to provide a better understanding of disability. Disability identification within housing-related data collections varies between programs and jurisdictions, information about disability group and severity is generally not captured, and disability information may be out of date

AT A GLANCE

Aim:

To understand whether an accurate indicator of disability can be derived from administrative data and used for reporting

What was found:

- The derived indicator performed well for individuals aged 25-64 years (prevalence estimates aligned with SDAC).
- The indicator underestimated disability prevalence for people aged 65 years and older. Including aged care data may help address this gap.
- The indicator underestimated disability for children and young people. More reliable ascertainment may be possible using education and health data.

What's Next:

- Validation of derived indicators through linkage to SDAC.
- Reporting on housing services and supports used by people with disability.

for some service recipients. The derived disability indicator can be applied consistently across housing data sources and can be used to track disability changes over time.

- Linked the derived disability indicators with the Australian Census of Population and Housing and SDAC data and commenced analysis and validation of the derived indicators.

In-scope population for this test case

- Residents of New South Wales, Queensland, South Australia or Victoria at any time during 2010–2020.
- All age groups.

KEY FINDING:

Approximately **2.5 million** people with disability were identified from linked administrative data for 2018-19. This compares to an SDAC estimate of about 3.7 million people for the included states (NSW, QLD, SA and VIC).

Notes: 2018-19 financial year was used to align as closely as possible with SDAC 2018.

6.1.1 Identification of people with disability and methodology

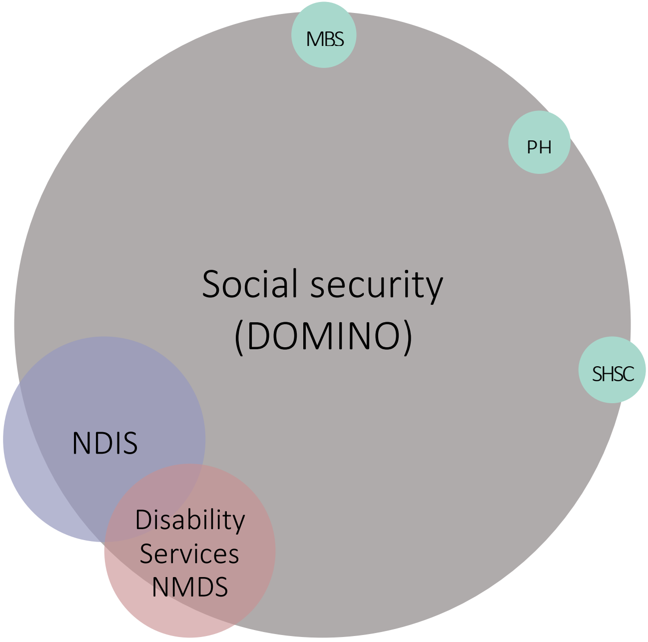
The following set of disability indicators were derived from the linked data:

- Disability (yes; no)
- Severe disability (severe or profound disability; other)
- Disability group (sensory or speech; intellectual or learning; physical; psychosocial; acquired brain injury; other), chosen to align as closely as possible with SDAC disability groups

<table>
<tr><th colspan="3">Three core data sets:</th></tr>
<tr><td>National Disability Insurance Scheme (NDIS)
All NDIS participants were counted as part of the disability cohort.</td><td>Social security data (DOMINO)
The following groups were counted as part of the disability cohort:
• Recipients of Disability Support Pension, Youth Disability Supplement, Mobility Allowance, or BSWAT Payment.
• People with long-term incapacity or partial capacity to work.
• Carer Payment or Carer Allowance recipients.
• People attending special schools.</td><td>Disability Services National Minimum Data Set (DS-NMDS)
DS-NMDS collated data from states and territories on specialist disability services provided under the historic National Disability Agreements. Data collection ceased in 2019, with transition of the disability services to the NDIS.
All recipients of disability services who appeared in DS-NMDS were counted as part of the disability cohort.</td></tr>
<tr><td colspan="3">Supplementary data sets:
• Medicare Benefits Schedule (MBS) – people in these data collections who received disability-specific services were included in the disability cohort.
• Pharmaceutical Benefits Scheme (PBS) – was used to identify potential psychosocial disability.
• Public Housing and State-Owned and Managed Indigenous Housing (PH & SOMIH), Specialist Homelessness Services Collection (SHSC) – people in these data collections identified as having disability were included in the disability cohort.</td></tr>
</table>

How did each data source contribute to the creation of the disability cohort?

Figure 6.1.1. Overlap between data sources used to identify the disability cohort



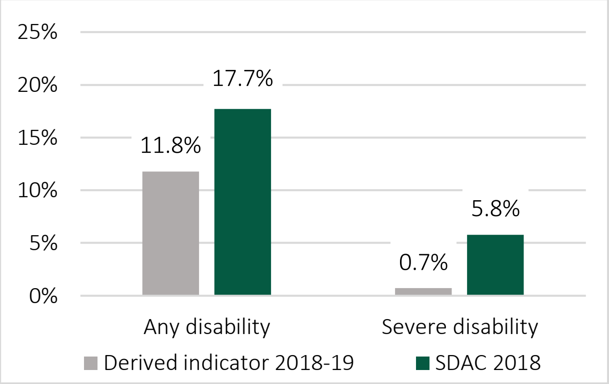
- Social security data (DOMINO) identified **98%** of people with disability in the 2018-19 cohort (**83%** were identified from DOMINO only).
- NDIS data identified just under **11%** of the cohort (1.1% were identified from NDIS data only).
- Disability Services NMDS identified less than **8%** of the cohort (0.3% were identified from DS-NMDS only).

Notes:
MBS = Medicare Benefits Schedule;
PH = Public Housing and State Owned and Managed Indigenous Housing;
SHSC = Specialist Homelessness Services Collection.

The placement of MBS, PH and SHSC in Fig. 1 does not indicate there was no overlap with NDIS and/or DS-NMDS.

How well does the derived indicator compare to population estimates?

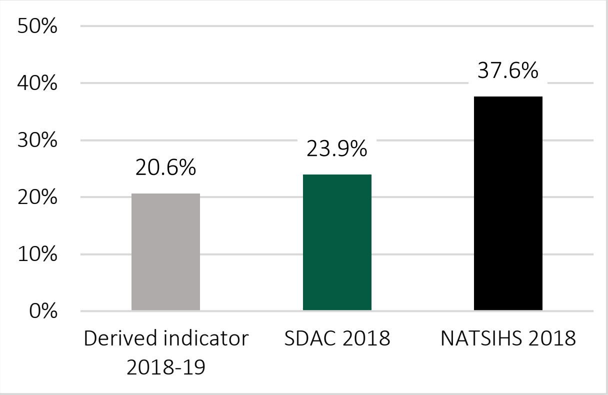
Figure 6.1.2. Overall disability prevalence



- The prevalence of disability was estimated at **11.8%** using the derived indicator compared to **17.7%** using SDAC.
- Derived indicator for **severe disability substantially underestimated** the prevalence of severe disability compared to SDAC.

How well does the derived indicator work for specific groups?

Figure 6.1.3. Disability prevalence among First Nations Australians

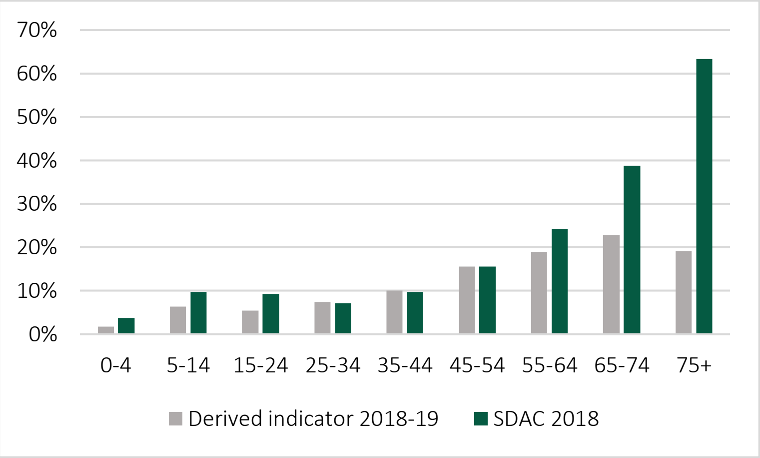


- The prevalence of disability among First Nations Australians was higher than among non-Indigenous people; **20.6%** using the derived indicator based on linked data, compared to **10.3%** for non-Indigenous people.
- The derived disability indicator aligns relatively well with the SDAC estimate (**20.6%** using the derived indicator vs **23.9%** using SDAC). Both are significantly lower than the NATSIHS estimate of **37.6%** (Fig. 6.1.3).

Notes:

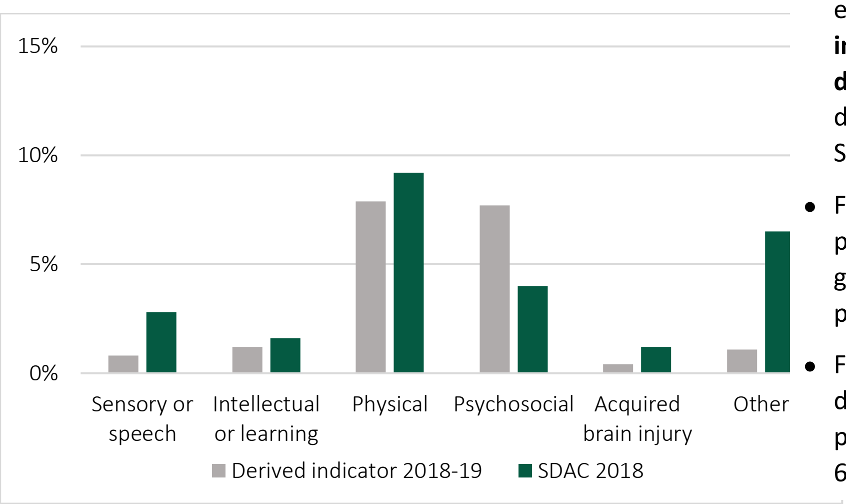
1. For SDAC 2018 and NATSIHS 2018, reported proportions are for all states and territories rather than the four states included in the test case.
2. NATSIHS uses a different (shorter) version of disability module compared to SDAC; the shorter module tends to produce higher estimates of disability compared to the SDAC module.

Figure 6.1.4. Disability prevalence, by age group



- For people **aged 25–64 years**, the prevalence of disability was similar to SDAC (**12.6%** using the derived indicator compared to **13.6%** using SDAC for that age group).
- The derived indicator worked particularly well for people aged 25-54 years (Fig. 6.1.4).
- For children (aged 0–14 years), young people (aged 15–24 years), and older people (aged 65 years and above), the **derived indicator substantially underestimated** the prevalence of disability compared to SDAC (Fig. 6.1.4).

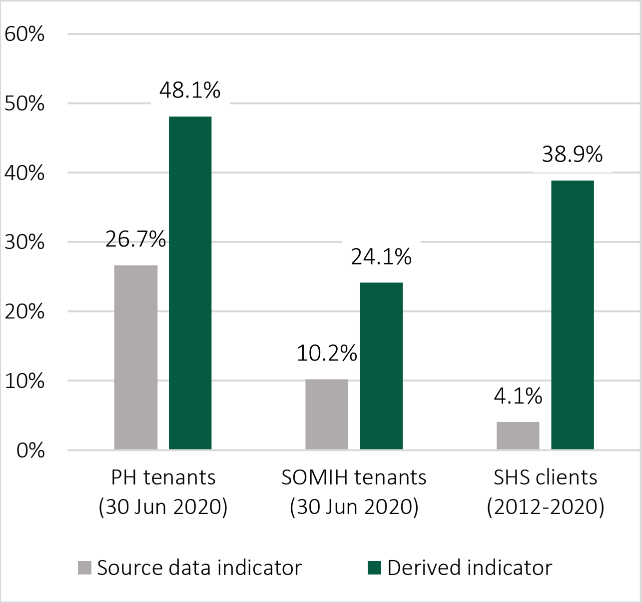
Figure 6.1.5. Disability prevalence by type, for people aged 25-64 years



- For people aged 25-64 years, the estimated prevalence of **intellectual/learning** and **physical disability** was **similar** using the derived disability group indicator compared to SDAC (Fig. 6.1.5).
- For **psychosocial disability**, estimated prevalence using the derived disability group indicator was **almost double** the prevalence in SDAC (Fig. 6.1.5).
- For all **other disability groups**, the derived indicator **underestimated** the prevalence compared to SDAC (Fig. 6.1.5).

What is the impact of using the derived indicator when reporting on housing supports?

Figure 6.1.6. Comparison of disability prevalence using derived and source data indicators



- Some housing services providers collect information about disability status of their clients.
- However, definitions of disability vary between programs and jurisdictions and disability information may be out of date for some service recipients.
- The derived indicator based on linked data increased the identification of people with disability about **two times** for the Public Housing and State-Owned and Managed Indigenous Housing data; and **nearly ten-fold** for the Specialist Homelessness Services Collection data (Fig. 6.1.6).

Notes:

PH = Public Housing; SOMIH = State Owned and Managed Indigenous Housing; SHS = Specialist Homelessness Services.

PH statistics are for NSW, QLD and SA only; tenants in Victoria were excluded as data relating to their disability status was not supplied by Victoria.

SOMIH statistics are for QLD and SA only; SOMIH data for NSW were not supplied for this test case; Victoria does not have a state-owned and managed Indigenous housing program.

6.1.2 Data gaps and areas for data improvement

- The derived indicator based on linked administrative data **underestimated disability** prevalence for **people aged 65 years and older**, as well as **children and young people up to age 24 years**. Inclusion of health data and aged care data could help address this problem.
- The group of people identified by the derived disability indicator is largely comprised of individuals who are NDIS participants or receive social security payments, and is thus a narrower and administratively-defined group compared with the SDAC disability population.
- Other data sources which could provide more information about disability are hospital data (including admissions and emergency department presentations) and mental health services data, utilising diagnoses that may be expected to be associated with disability.
- Challenges in achieving a national picture from services data collected at state (or local) level arise if there are **inconsistencies** in data coverage and completeness, or in data definitions across different jurisdictions.
    - National Minimum Data Sets (NMDS) were used extensively in this test case. These data sources provide a valuable source of nationally collated state data but were found in some cases to have limitations which users of the data need to be aware of. The quality and consistency of nationally collated state data will need to continue to be an area for investment and data improvement for the enduring NDDA.
- The derived **severity indicator** significantly underestimates the prevalence of severe or profound disability.
    - Data items in social security data (DOMINO) relating to results of assessments against impairment tables (not provided for this test case) may be used to derive severity information for Disability Support Payment (DSP) applicants. However, there is no severity information in DOMINO for non-DSP recipients, which may leave gaps for significant proportions of the disability cohort identified via DOMINO.
    - Where present, severity indicators in administrative data (e.g., NDIS) do not align with the SDAC measures of severe or profound disability.
- Identifying **First Nations Australians** with disability – existing administrative data may inaccurately reflect disability for several reasons:
    - No word for 'disability' in traditional languages – the Aboriginal and Torres Strait Islander communities talk about individuals in enabling, not deficit terms.
    - People may not want to identify as Aboriginal or Torres Strait Islander to a government services provider.
    - The current lack of services in remote and very remote areas results in unmet need, as well as underestimation of First Nations Australians with disability.

  - Lack of culturally appropriate engagement with people seeking help may result in disengagement from services, and non-identification.
  - Not asking the right questions (or not using the right language) at service entry point may result in 'hidden' disability, such as inadequate housing accessibility supports if the lead tenant does not identify anyone in their household as having disability.
  - ➔ Research led by the First Nations Australians is needed to address this data gap.
- The group of people identified by a disability indicator derived from linked administrative data will be affected by:
  - the use of different definitions of disability in different administrative data sources, which align to a greater or lesser extent with key population data sources such as 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 (which may vary over time with changes in policy);
  - the varying frequency of collecting or reporting disability data.

These factors should be taken into account when considering the validity and utility of a derived disability indicator for a specific purpose and caveats that may be needed to guide interpretation of resulting data.

The age limitations of the derived indicator will be important to consider when using the indicator for specific cohorts of people such as tenants in public housing (who tend to be either very young or old), or those in receipt of specialist homelessness services (who tend to be young). It is also an issue when reporting about Aboriginal and Torres Strait Islander populations which tend to have a relatively young age structure.

Next steps

- The comparison against benchmarks was the first step in testing the derived indicator. Individual-level linkage to SDAC and Census 2016 will provide further insight into validity of the derived disability indicators.
- Reporting on housing services and supports used by people with disability, disaggregated by demographic and disability groups, and compared to people without disability.

6.2 Indicators of psychosocial disability in Victoria

This test case (Mental Health test case), led by Victoria, identifies cohorts of people living with disability and mental health issues (i.e. psychosocial disability with and without other co-existing disabilities, and disabilities with mental issues) and examines their characteristics, use of services and supports, and service gaps. It aims to assess which supports are effective in improving outcomes for people living with disability and mental health issues.

A list of datasets included in this test case is provided in the Appendix.

Highlights and key accomplishments since June 2021

- Engaged and onboarded analysts and academic research partners on the test case analytic team (August 2021).
- Finalized transfer of the complete linked dataset to the analytic environment (August 2021). Data consist of:
  - Disability service user populations: recipients of
    - Disability Support Pension payments;
    - National Disability Agreement funded services;
    - NDIS applicants
  - Mental health service users: individuals with records in
    - Public clinical mental health dataset (CMI/ODS);
    - Victorian Admitted Episodes Dataset – people admitted to private hospital mental health beds;
    - Victorian Emergency Minimum Dataset– people presented to Victorian public hospital emergency departments for mental health reasons;
    - Medicare Benefits Schedule (MBS) data;
    - Pharmaceutical Benefits Scheme (PBS) data.
- Conducted preliminary investigations of the most appropriate methodologies to identify disability (including psychosocial disability) and mental health need.

Noting the recency of data access, findings for this test case are preliminary only and subject to change and further validation.

AT A GLANCE

Aim:

To identify utilisation of services and supports by cohorts of individuals living with disability and mental health issues in Victoria

What was found:

- A preliminary disability indicator indicates that disability service user populations have increased since 2008.
- 57.6% of mental health service users presented to a Victorian emergency department for a mental health reason. This may be an underestimate, as the public mental health system only represents a small proportion of mental health service use.

What's Next:

- Review and refinement of disability and mental health cohort definitions, including validation of a definition for psychosocial disability.
- Descriptive analysis of service utilisation.

6.2.1 Identification of people with disability and methodology

The linkage approach has imposed some criteria on the definition of disability service users. Data has been linked for those clients who:

- had a valid Medicare Consumer Directory record;
- received disability related supports either through:
  - Disability Support Pension (DSP) at any time between July 2008 – April 2020;
  - National Disability Agreement (NDA) funded service at any time between July 2008 – June 2019;
  - or applied for the NDIS between July 2013 – May 2020;
- and had Victorian residency at any time in any of these datasets.

The following figures provide a summary of the initial disability service user population and their key characteristics based on these criteria. They report the number of people living in Victoria eligible for DSP[5], using NDA-funded services and active NDIS participants with an approved plan.

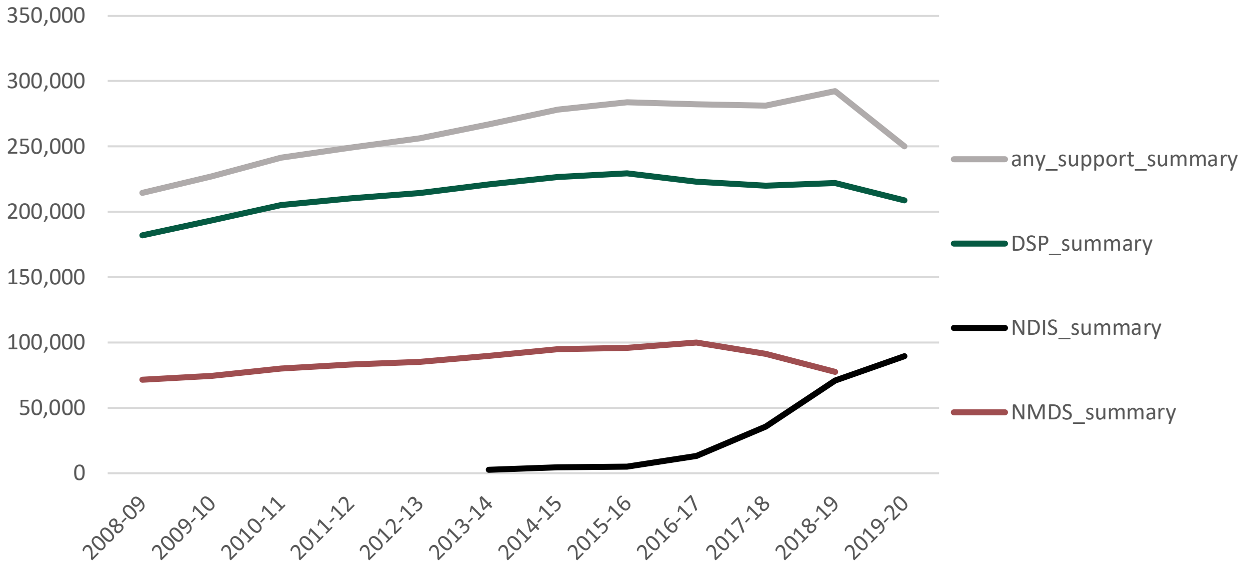
In-scope population for this test case

- Residents of Victoria receiving disability support services at any time between 1 July 2008 and 30 June 2020.
- All age groups.

---

[5] These numbers are an over-count, as the official statistics for DSP only count people at the time they receive payments. This can take months after initial application – the point in time the person is eligible.

Figure 6.2.1 Disability service users by service type



* June 2019 marks the end of NDA funded services.
** July 2013 marks the start of NDIS supports.

Figure 6.2.2 Disability service users by gender

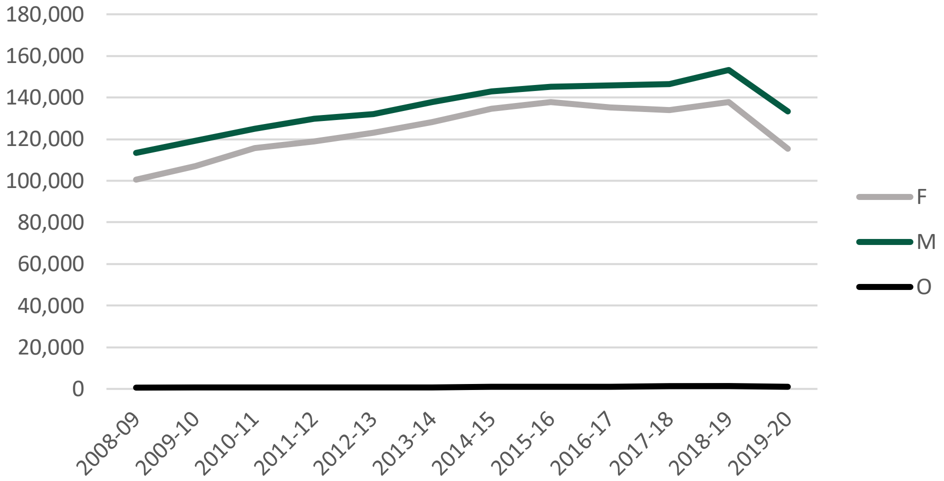


Figure 6.2.3 Disability service users by culturally and linguistically diverse status

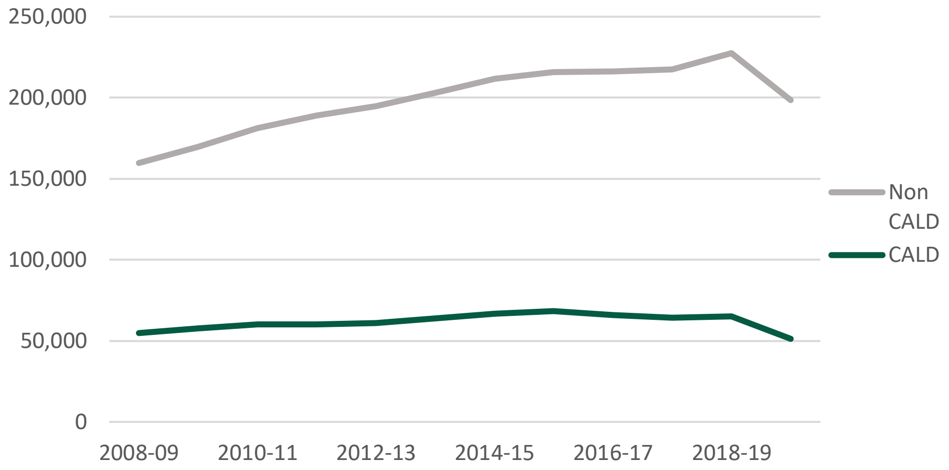


Figure 6.2.4 Disability service users by location

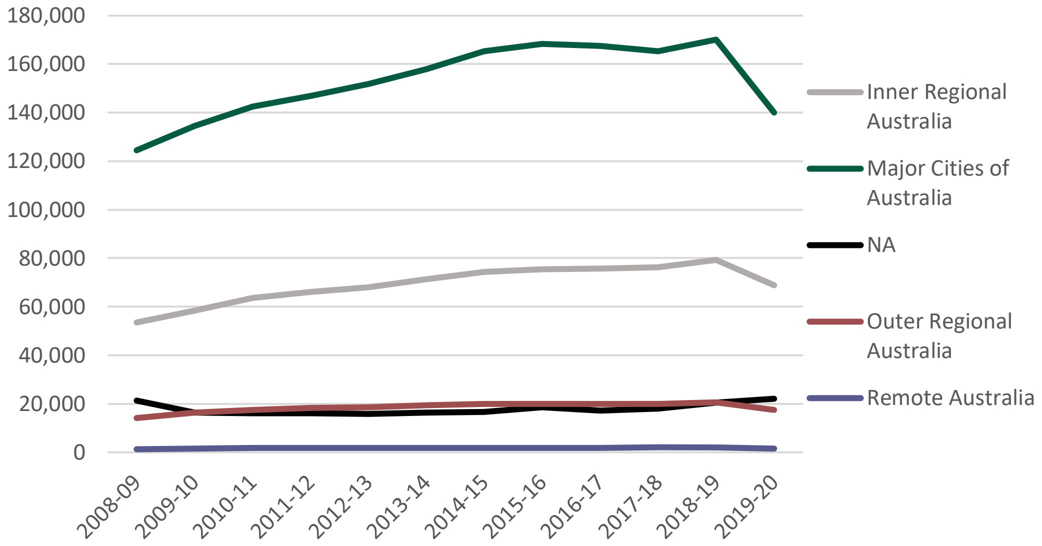
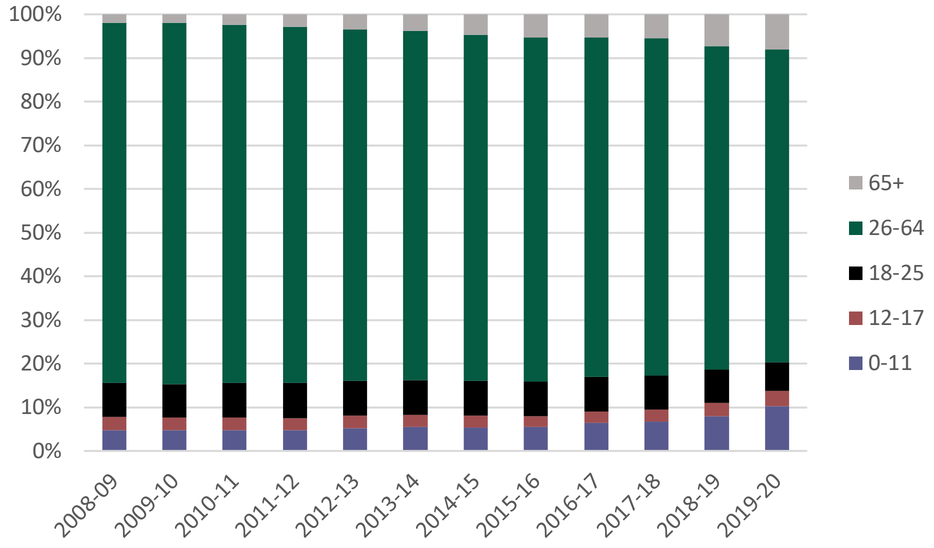


Figure 6.2.5 Disability service users by age group



Mental health service users are defined primarily as clients who access the state based public clinical mental health system. Other datasets are also used to expand this group, including people who present to Victorian public hospital emergency departments for mental health reasons (VEMD) and people admitted to private hospital mental health beds (VAED).

It is recognized that this cohort represents only a small portion of mental health service users, as the public mental health system only sees around 1 per cent of this Victorian population, while epidemiological studies suggest it should be at around 3 per cent (RCVMHS, 2021). The cohort is further expanded by adding clients who access mental health services via their GPs (Medicare Benefits Scheme items) and access mental health related items through the Pharmaceutical Benefits Scheme.

Table 6.2.1 presents the number of mental health service users in the disability population by service system.

Table 6.2.1 Mental health service users (of disability service user population) FY2009 – FY2018[1]

Dataset component	Dataset source	FY2009	FY2010	FY2011	FY2012	FY2013	FY2014	FY2015	FY2016	FY2017	FY2018
Public specialist active MH client	CMI/ODS	27,599	29,263	30,013	29,337	31,307	31,662	32,202	32,299	31,441	32,956
Private MH clients	VAED	1,612	1,968	2,180	2,203	2,219	2,254	2,442	2,534	2,611	2,601
Presented to Victorian Emergency Departments for MH reason,	VEMD	14,436	14,762	15,297	16,518	17,362	17,515	17,683	18,254	18,166	18,296
Clients who received MH related MBS service	MBS	14,692	16,516	19,958	22,474	26,415	29,555	32,642	37,165	42,115	46,206
Clients who received MH related PBS item	PBS	172,501	179,472	185,488	198,236	209,624	213,652	214,927	217,477	214,829	217,611

Notes:

[1] Preliminary results only, subject to change as further refinements will be implemented.

[2] This table counts the number of active clients of the public mental health system (were admitted or received a contact in the time period).

[3] This table counts the number of VAED private patients with care type 5

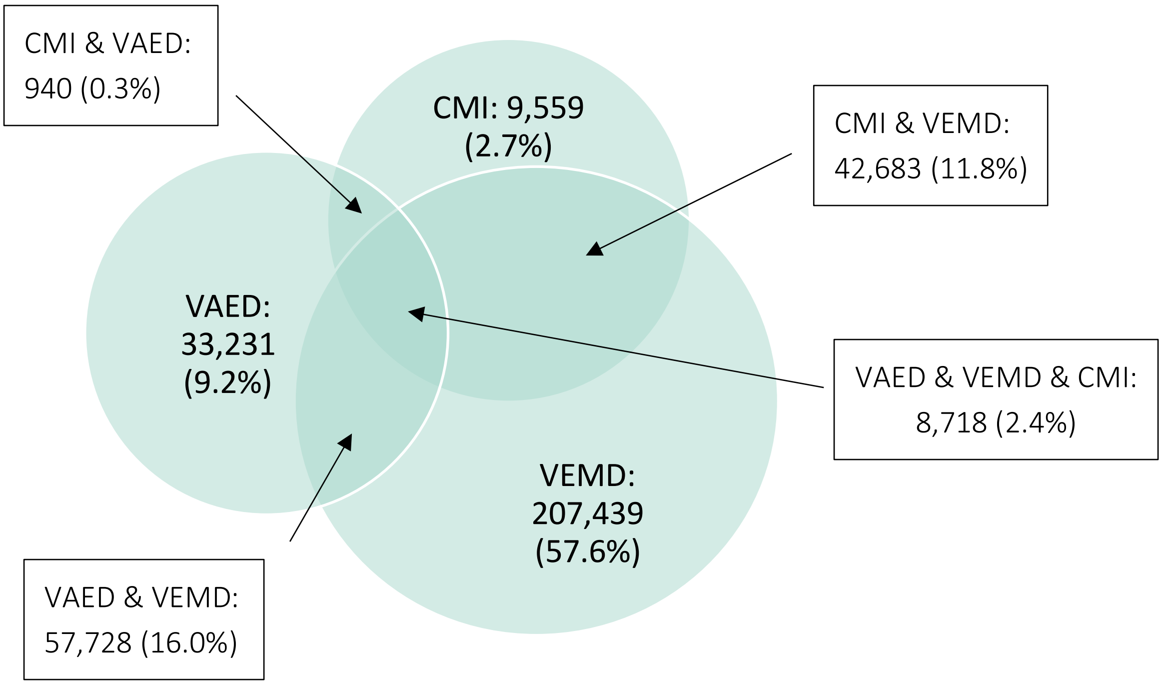
[4] This table counts the number of VEMD patients with indicators: 1) referred by code 16/18/21; 2) human intent code of 2/18/19/20; 3) diagnosis code F01-F99/Z004/Z046/Z915/R4581; 4) seen by mental health practitioner date/time not empty; 5) departure status code 17/23/25/26/31. The counts excludes type of visit COVID-19 and Triage cat.

[5] This table counts patients who accessed mental health related MBS services and had a postcode in Victoria for the MBS service.

[6] This table counts patients who accessed mental health related PBS items and had postcode in Victoria for this PBS item.

[7] The Victorian postcode requirement has only been applied to the MBS and PBS datasets at this stage.

Figure 6.2.6 Venn diagram of disability and mental health service users in the Victorian mental health datasets FY 2016 – FY 2020 (preliminary findings) [1]



[1] These numbers are subject to change, as the definition of mental health users is being reviewed by the Team.

VAED = Victorian Admitted Episodes Dataset; VEMD = Victorian Emergency Minimum Dataset; CMI = public clinical mental health dataset

6.2.2 Data gaps and areas for data improvement

The dataset for this test case includes all Commonwealth and state data for the disability service user population (see Highlights and Key Accomplishments above) and gives an opportunity to significantly expand the disability and mental health cohorts.

Currently, the mental health user cohort includes Victorians in the disability service user population who:

- were a client of the Victorian public specialist system; or
- were admitted to a Victorian private mental health bed; or
- presented to a Victorian public emergency department for a mental health reason; or
- received mental health related MBS services; or
- were dispensed a mental health related PBS item.

Information that is not captured in these datasets but would be beneficial for the identification of people with mental health issues include:

- clients who access mental health helplines;
- consultations with work/school-based practitioners not included in the MBS;
- consultations private practitioners not included on the MBS;
- private hospital emergency department mental health related presentations;
- drugs dispensed within the hospital system not entered on PBS.

The dataset does not capture potential users of mental health services, as this data is currently not readily available.

Data quality issues will be continuously documented and managed through the project.

Next steps

- Review and refinement of definitions and rules around the cohorts and their key characteristics, including:
    - Review of the disability definition among the disability service user population;
    - Review of the mental health service use definition across VAED, VEMD, CMI, PBS, MBS, NDIS;
    - Development and validation of a definition for psychosocial disability, based on a combination of data items from the CMI/ODS dataset;
    - Agreement on the most suitable approach to obtain reliable demographic data (including culturally and linguistically diverse and Aboriginal and Torres Strait Islander status, gender, location)
- Development and verification of the mental health service user cohort and classification into broad mental health and disability categories. These may include categories such as: ongoing intensive public specialist; ongoing public specialist; short term public specialist; primary care only; and psychosocial / physical disability categories.
- Descriptive analyses by service user categories.

6.3 Indicators of disability and developmental vulnerability among children in NSW

The NSW-led test case *Indicators of disability and developmental vulnerability among children in NSW* (Early Childhood test case) explores supports through mainstream and targeted services across state and national health, education and human services systems that promote optimal educational outcomes. It aims to understand the relationship between supports for children with disability and developmental delay and a range of developmental and educational outcomes, including school participation and progression.

A list of datasets included in this test case is provided in the Appendix.

Highlights and key accomplishments since June 2021

The test case population is represented in 23 Commonwealth and NSW datasets providing information on mainstream health, education, social, and targeted disability supports and services (details listed in the Appendix). Some mainstream support services include GP visits, hospital admission and enrolment in education services.

- The test case population comprises **2.33 million** children born between 2003-2019.
- Among this cohort, **10.7%** of children were identified as living with a disability by age 5 years.
- By age 7 years, **13% of children were identified as having disability**. More than 25% of children identified with disability had multiple disability types. A common complexity is the combination of intellectual/learning and psychosocial disabilities.
- Disability was most commonly first identified in the MBS dataset, which identified around **66% of the children with disability in the Early Childhood test case population**.
- As shown in Figure 6.3.1 below, **more than 99% of children identified with disability received at least one mainstream or targeted support service**, 48% received targeted disability supports and more than 80% of children identified with disability were enrolled in Early Childhood Education.

AT A GLANCE

Aim:

To understand the relationship between supports and outcomes for children with disability and developmental delay

What was found:

- 10.7% of children were identified as living with a disability by age 5, increasing to 13% by age 7.
- More than 99% of children with disability received support.
- Disability was most commonly identified from Medicare Benefit Scheme data.

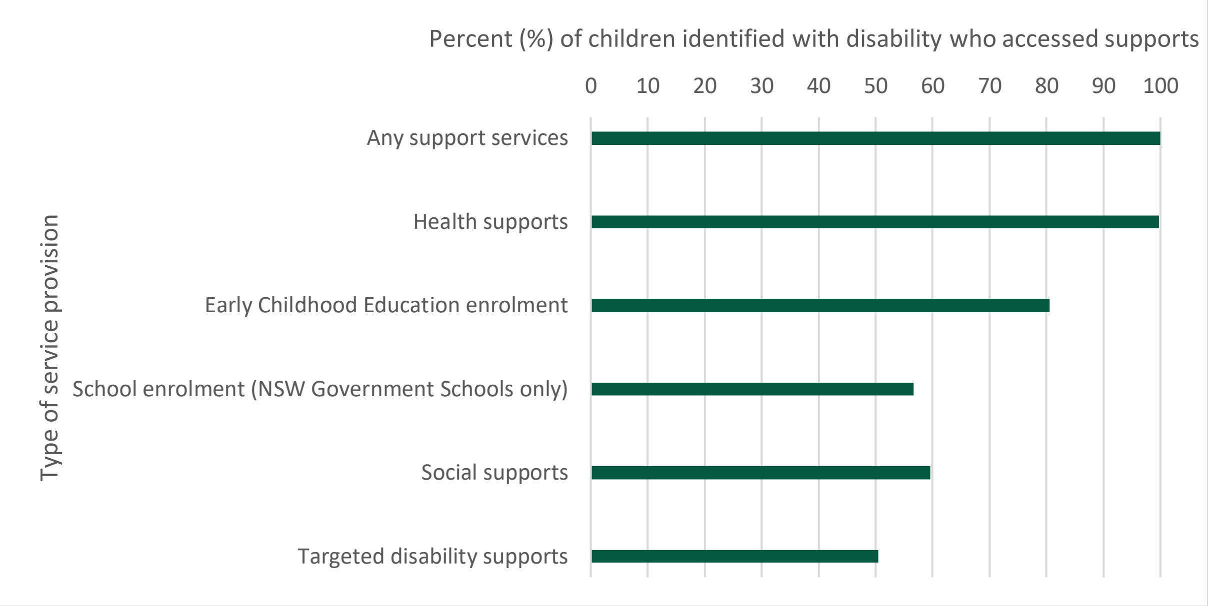
What's Next:

- Assessment of:
  - impact of disability on school attendance;
  - impact of supports on outcomes;
  - factors impacting access to supports.

In-scope population for this test case:

All children in NSW born between 2003-2019

Figure 6.3.1. Proportion (%) of children identified with disability who received different types of support services.



6.3.1 Identification of people with disability and methodology

Indicators of Disability and Developmental Vulnerabilities

Linkage of several datasets enabled the identification of groups of interest for targeting early childhood supports and services. Ten datasets (see Table 6.3.1) were used to determine three sub-cohorts of children with Disability or Developmental Vulnerabilities as follows:

- **Any Disability**: 306,982 (13.2%) children were identified as having disability (medically verified or non-medically verified), received targeted disability support or adjustments in school. They were identified using information from all of the datasets used to define disability (see Table 6.3.1);

- **Medically Verified Disability**: 258,183 (11.1%) children were identified as having disability using NDIS, MBS and DS-NMDS. These are datasets that require medical verification of disability.
- **Developmental Vulnerabilities**: An additional cohort of interest for targeted early supports identified 45,147 (9.8% of eligible children among 462,067 AEDC records) children who scored below the national 10th percentile on two or more domains of the Australian Early Development Census (AEDC)[6]. Children with developmental vulnerabilities may also be identified in the above two sub-cohorts.

Table 6.3.1. Datasets utilised to determine disability and developmental vulnerabilities for the NSW Early Childhood test case.

Commonwealth or National Record Collections	1. National Disability Insurance Scheme (NDIS) including Early Childhood Early Intervention pathway 2. Medicare Benefits Schedule (MBS) 3. Australian Early Development Census (AEDC) 4. Disability Services-National Minimum Data Set (DS-NMDS)
State Record Collections	5. Nationally Consistent Collection Data on School Students with Disability (NCCD) 6. NSW Disability and Inclusion Program – Higher Learning Support Needs (DIP-HLSN) 7. NSW Preschool Disability Support Program (PDSP) 8. NSW DoE Student Disability Data collection: Integrated Funding Support (ESD-IF) and Support Class (ESD-SC) 9. NSW Annual (Community) Preschool Census (APC) 10. NSW Mid-Year Census: Government Preschools (MYC-GP) and Early Intervention (MYC-EI)

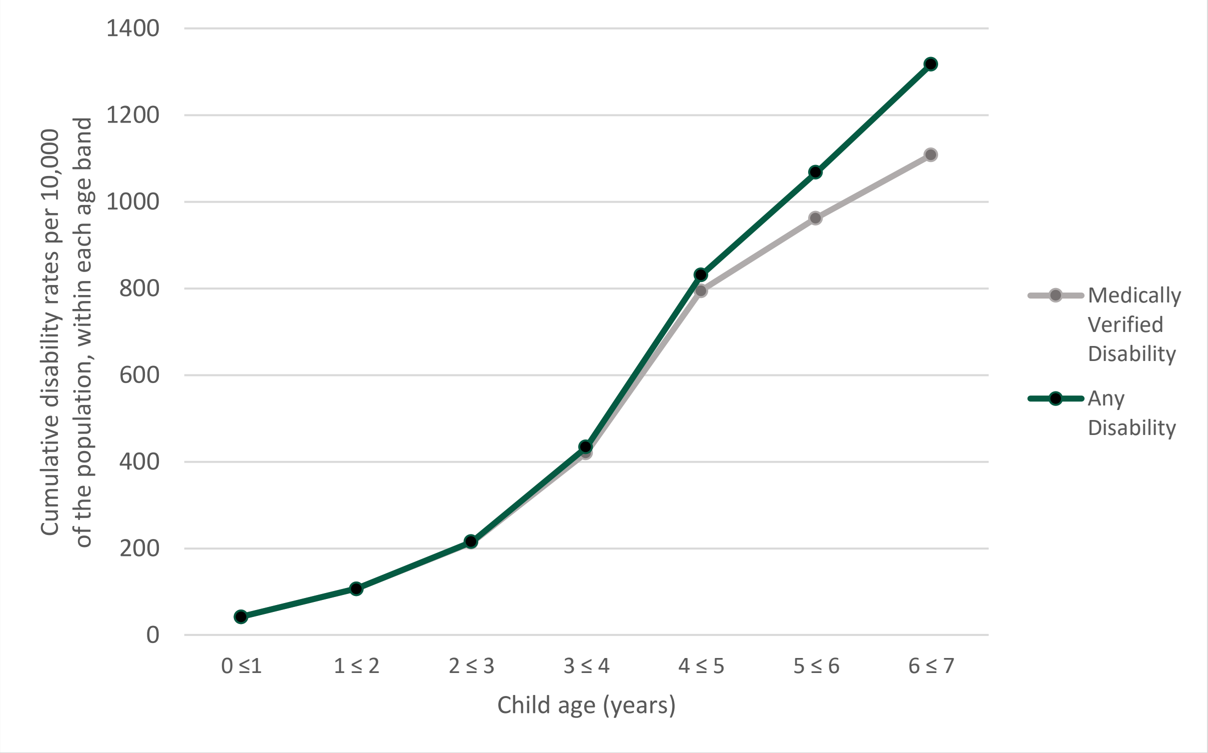
For children with any disability identified before the age of 7, ~13% of these children were identified in the NDIS, and around ~20% identified in NCCD, whereas ~73% were identified through the MBS. The contributions to identification in this test case are governed by policy changes and data availability (e.g., the NDIS is a recent scheme). Touchpoints through mainstream or targeted services captured in MBS are important for this age cohort.

The prevalence of disability identified by age 7 years in this test case is marginally higher than typically reported, where the prevalence estimates are often obtained from the Survey of Disability, Ageing and Carers (SDAC), for example, 3.7% for children aged 0-4 years, and 9.6% for young people aged 5-14 years in 2018. Figure 6.3.2 shows the cumulative rate of disability identification for both Any Disability and Medically Verified Disability sub-cohorts in this Test Case. The broad age band for SDAC and the differences in obtaining prevalence estimates from survey compared to large administrative data sources makes these estimates difficult to compare directly.

---

[6] The cut-off for an AEDC score to represent 'vulnerable' is based on the baseline set in the 2009 AEDC data collection. In 2009, children who scored below the 10th percentile (in the lowest 10 per cent) of the national population were classified as 'Vulnerable'. www.aedc.gov.au/dataguide.

Figure 6.3.2. Cumulative identification of Any Disability and Medically Verified Disability from birth to 7 years.



Types of disability. Disability was classified into 5 broad subtypes. Of the 306,982 children identified with a disability:

- 14.9% had a disability characterised as 'Physical/Diverse'
- 55.0% had a disability characterised as 'Intellectual/Learning'
- 18.5% had a disability characterised as 'Sensory/Speech'
- 39.6% had a disability characterised as 'Psychosocial'
- 6.9% had a disability characterised as 'Other Disability'

These groupings are not mutually exclusive. 26.7% of children identified with a disability (<7 years) had more than a single type of disability and of which, one third of them have both intellectual/learning disability and psychosocial disability.

The way the data was initially captured could limit the information around the number of disabilities recorded. The datasets used to identify disability limited the number of disabilities that could be recorded for each child, and the number of disability subtypes that were recorded. Medically verified disability was able to be categorised into 19 subtypes, while information from other (e.g., education) datasets was categorised into 5 disability subtypes.

Changes in the way the data is collected over time also impact on identification. In this test case the rates of intellectual disability may be an underestimate due to a change in MBS data collection 2010-2016 (see focus box in data improvement).

Indicators of Health, Education, Social, and targeted Disability support services

A range of mainstream health, education and social supports, as well as targeted disability supports accessed by the children with Any Disability are shown in Table 6.3.2. This table indicates the number and proportion of these children who had accessed mainstream and disability-targeted supports and services by age 7 years, and the service areas from which these supports were provided.

Mainstream and targeted supports delivered through **health service systems** were almost universal with 99.7% of children with disability aged 0-7 accessing services, of note over 99% accessed a GP (MBS provided) service.

Services and supports from **education service systems** are accessed at an older age (e.g. enrolment in early childhood education and school). As this cohort information refers to children aged 0-7 years, the proportion of children receiving supports through education service systems will not be comparable with other sector supports accessible from birth. Given this, government supported education service systems are an important source of mainstream support for children with any disability with over 80% accessing early childhood education services.

A range of support types are accessed through **social service systems** including childcare funding support to attend day care services without formal early childhood education programs, and out of hours care, through to child protection and out-of-home care. Over a quarter of children with any disability had some contact with the child protection system.

Around half the cohort accessed **targeted disability supports and services** through disability-specific and education service systems. Note that in this test case the health supports have not been separated out into targeted and mainstream supports as additional data and development work is required for this. The proportion of this cohort accessing **targeted disability supports** through education service systems will also be impacted by the age the disability was identified (Figure 6.3.2), where the support was provided (e.g. at school) and the introduction of support schemes (e.g. NDIS was only available from 2013, and only fully funded in NSW since 2019).

Table 6.3.2. Number (%) of children with Any Disability (< 7 years of age) who received supports.

Supports	N (%)
Any support services All of the below datasets	306,945 (99.9%)
Health supports Any MBS services (including GP and mental health services); NSW Hospital admission; NSW Emergency department presentation	306,184 (99.7%)
Early childhood education enrolment Centre-based day care (CCS/CCB); Preschool attendance (NSW Annual (Community) preschool census and NSW Mid-Year Census: Government preschools) *Note only pre-school age children are eligible for enrolment and attendance some of the children in the cohort are younger than pre-school age*	247,423 (80.6%)
Government School enrolment NSW School enrolment *Note only school age children are eligible to access this service and many children in the cohort are younger than school age*	173,918 (56.7%)
Social supports (child protection services) Contact with NSW Department of Communities and Justice (child protection services, including out-of-home-care)	79,548 (25.9%)
Social supports (other types of child-care) Family day care, Out of school hours care, and In-home care (CCS/CCB)	142,612 (46.5%)
Targeted disability supports Including: NSW Preschool Disability Support Program; NSW Disability and Inclusion Program-Higher Learning Support Needs; NSW Mid-Year Census: Early Intervention; NSW Department of Education Student Disability (Integrated Funding Support and Support Class); Nationally Consistent Collection Data on School Students with Disability; NDIS plan/Early Childhood Early Intervention pathway; Disability Services National Minimum Dataset	147,364 (48.0%)

6.3.2 Data gaps and areas for data improvement

Representativeness of the cohort

- While most data were available for the NSW child population, MBS data for the NSW test case was provided for a sub-cohort of the NSW population who met one of 15 criteria for increased risk of developmental vulnerability[7]. This means that there is limited capacity to compare mainstream health care use with that for the population at risk of vulnerability.
- Whole-of-population samples would ensure a more robust platform to inform policy.

---

[7] Vulnerability was determined according to factors identified in the linked records, as described in the "NDDA Early Childhood Supports Test Case: Linkage Report". This is distinct from developmental vulnerability identified in AEDC.

Additional data selection

- The inclusion of diagnostic codes in hospital admissions is likely to improve health service use patterns, and potentially the identification of disability.
- Information on health and social services and supports could be expanded to include community-controlled and child health services, community mental health, outpatient, Pharmaceutical Benefits Scheme data (for indicative medication), and additional information on medical services to delineate mainstream and targeted services. Additional information on social services, particularly child protection.
- The selection of variables could be expanded to provide more comprehensive information about familial risk and protective factors (e.g., mother's age at birth of child and targeted family intervention services, disability family support services, and Child Wellbeing Unit data).

Longer term data improvement

- Greater information on family, home and community environments would assist to understand the impact of services and supports in context, and to design more impactful wrap-around services.
- Additional factors derived via interview, such as community connections, culturally and linguistically diverse characteristics, should be considered in future data collections.
- A need for better quality meta-data and data dictionaries was highlighted in this test case. A case study of a data quality issue identified during analysis is shown below.

Case study: MBS Health Assessments

The MBS Health Assessment items 701, 703, 705, 707 are currently assigned for health assessments undertaken on people who meet set criteria related to risk of chronic disease, one of these criteria is intellectual disability. However, historical changes in the 'Healthy Kids Check' service noted between May 2010 and April 2016, meant routine 4 year-old health checks were captured in these assessment items. This resulted in nearly all children aged 4 years being captured and identified with intellectual disability for a period of 6 years. For this test case, for this period these health assessments could not be used to identify children with intellectual disability. We cannot be sure as to whether the rates of intellectual disability are accurate owing to these changes in the use of MBS codes during the study period. This illustrates how changes in data collection and or policy can impact on the identification of disability and prevalence reporting.

Next steps

- The impact of disability and developmental vulnerability of attendance at education services
- The impact of certain services and supports on education outcomes of children identified with Any Disability, Medically Verified Disability, and/or Developmental Vulnerabilities, e.g. the impact of early provision of services and supports on childhood developmental vulnerability at school entry; the impact of a range of early childhood supports and services on educational attainment (e.g., NAPLAN) in early and middle childhood for children with and without disability.
- Factors impacting the access of services and supports for children with disability and developmental vulnerability.

6.4 Indicators of disability in the NSW criminal justice system

The test case *Indicators of disability in the NSW criminal justice system* (Justice test case), co-led by DSS and NSW, examines the victimisation and offending rates of a cohort of individuals with intellectual, physical, sensory and psychiatric disabilities. It aims to examine factors associated with increased risk of contact with the criminal justice system and understand the use of diversionary options for offenders with disability.

A list of datasets included in this test case is provided in the Appendix.

Highlights and key accomplishments since June 2021

- Approximately **843,800** people within the test case cohort were identified as having **disability** at any time during 2009-2018.
    - Of these, 230,200 people were identified as having cognitive disability, 415,000 physical disability, and 477,200 psychosocial disability.[8]
- Overall, **332,285** people were identified as having **severe disability** at any time during 2009-2018.
- Defined six study cohorts with varied forms of contact with the NSW criminal justice system, and generated summary statistics for each cohort using multiple measures of disability.
- Developed a First Nations indicator in line with documented best practice and consultations with the newly established Aboriginal Perspectives Expert Panel.
- Completed data curation and development work to enable linkage across NSW Police Force victim records to provide a longitudinal view of victimisation.

6.4.1 Identification of people with disability and methodology

People with disability may not be identified when in contact with the justice system. Within this test case the disability cohort was drawn from the linked administrative data to include:

- Original Disability Indicator: people who received a **core disability support or service** – these included NDIS; Disability services provided under historic National Disability Agreements; and Disability Support Pension; and

AT A GLANCE

Aim:

To examine victimisation and offending rates among individuals with disability and understand factors associated with increased risk of contact with the justice system

What was found:

- 843,800 individuals were identified as having disability.
- Of individuals in receipt of a core disability service or support, 28% had contact with the justice system as a victim and 16% as an offender.
- These proportions increased when people who were not in receipt of a core disability service or support were identified as having disability in the linked data across health, housing and social services.

What's Next:

- Assessment of factors associated with contact with the justice system.
- Exploration of outcomes of charges involving people with disability.

---

[8] A person could have multiple types of disabilities identified; or no specific disability identified.

- Expanded Disability Indicator: people who may not be in receipt of a core disability support or service but were identified as having disability in the linked data across health, housing and social services based on their diagnoses or services received.

In-scope population for this test case

- Residents of NSW at any time during 2009-2018.
- Aged 10 years or older.
- In contact with criminal justice system as victim or offender, **OR** recipients of **core** disability services and supports

Table 6.4.1. Datasets used to define disability in the Justice test case

Data source	Identifier of disability	Disability sub-groups based on:
National Disability Insurance Scheme (NDIS)	**All persons accessing NDIS** All persons with conditions indicative of disability	Participant disability ICD codes
Disability Services National Minimum Data Set (DS-NMDS)	**All recipients of disability services**[9]	Primary and other significant disability codes
Social security (DOMINO)	**All Disability Support Pension (DSP) recipients** AND All social security payment recipients and individuals with selected permanent medical conditions[10]	Broad medical condition codes[11]
Medicare Benefits Schedule (MBS)	Persons receiving disability-specific medical services	MBS Item numbers

---

[9] Excluding services relating to research and evaluation, training and development, peak bodies.

[10] Medical codes in DOMINO were provided for the NDDA Pilot as broad groupings only. Broad medical groups were classified as disability if more than 50 per cent of specific conditions within that group were indicative of disability. This may result in some false positives and false negatives in identifications of disability.

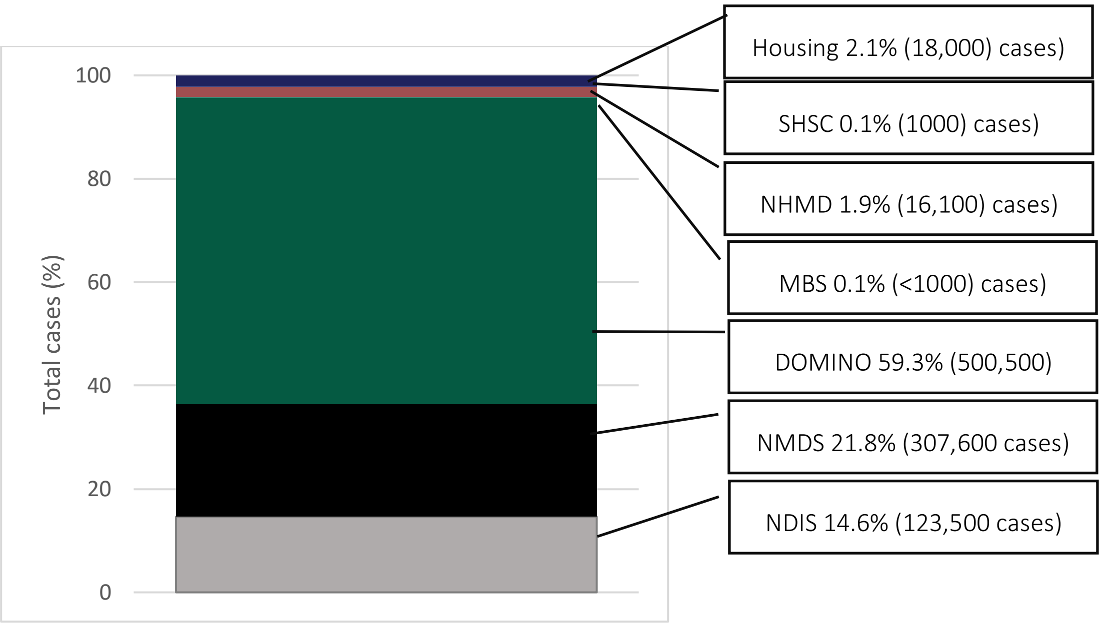
[11] Ibid.

National Hospital Morbidity Database (NHMD)	Persons with diagnoses indicative of disability	Primary and additional diagnoses based on ICD codes
Specialist Homelessness Services Collection (SHSC)	All people identified in SHSC data are considered as needing disability services	n/a
Public Housing (PH)	All people identified in PH data as having disability	Recorded disability groups
Specific to justice offending cohort		
NSW Reoffending Database (ROD)	All offenders who ever had mental health outcome for any offence AND All people in custody referred to Statewide Disability Services (SDS) AND All people with IQ < 70	Disability codes recorded in SDS referral

How did each data source contribute to the creation of the disability cohort?

The cohort creation process added data sources sequentially. Additional persons with disability identified by each new data set are shown in the figure below. More than half of identified cases were indicated from a diagnosis recorded in DOMINO. It seems unlikely that this many cases would be undetected in other datasets, therefore it is probable that the inclusion of all cases from use of higher order categories was too inclusive.

Figure 6.4.1. Contribution of data sources used to define the cohort



Indicator of severity

Administrative data sets in general lack indicators of severity of disability. Where present, these may not be consistent across data sets, or comparable with external data sources (such as surveys). In the justice context complexity (presence of several different disabilities) may be more reliably assessed than severity. However, higher complexity does not always mean higher severity.

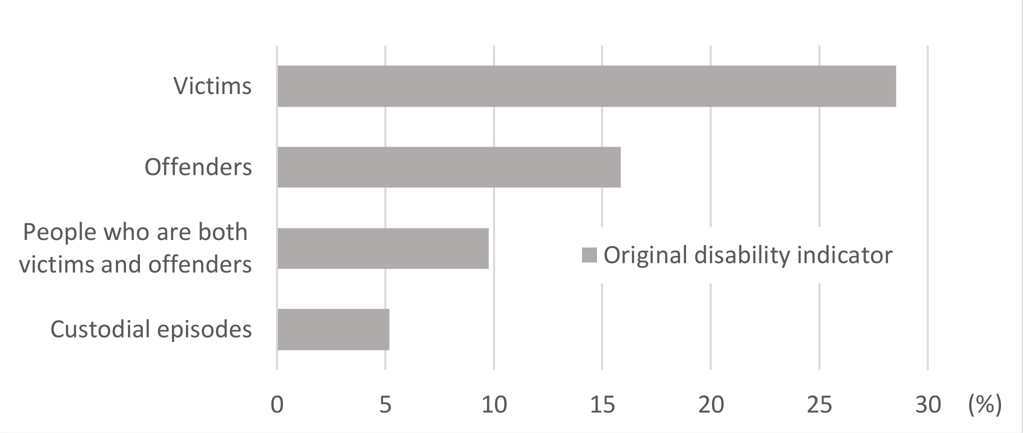
Table 6.4.2. Indicators of severity in the linked data for the Justice test case

Data source	Methodology for severity indicator
National Disability Insurance Scheme (NDIS)	• NDIS participants with severity score indicating 'Low function' • Recorded in NDIS as young persons in residential care
Disability Services National Minimum Data Set (DS-NMDS)	• DS-NMDS services recipients who always or sometimes needed help and/or supervision with a range of life activities • Disability services recipients whose informal carer assisted with activities of daily living: self-care, mobility, or communication
Social security (DOMINO)	• Individuals with current work capacity of 0 hours per week.
Medicare Benefits Schedule (MBS) National Hospital Morbidity Database (NHMD) Specialist Homelessness Services Collection (SHSC) Public Housing (PH)	Not used -no suitable severity indicators
NSW Reoffending Database (ROD)	• Severe intellectual disability recorded for individuals with IQ < 35

Interaction of people with disability with Justice system

Inclusion in this test case was based on a person having a contact with the NSW criminal justice system, or being in receipt of one of the three core disability services or supports (NDIS, DS-NMDS, DSP) in the 10-year period from 2009-2018. Beyond the cohort of people with a disability, five additional non-mutually exclusive cohorts representing different forms of contact with the criminal justice system in the study period were identified for analysis. These include young and adult offenders with a criminal court appearance, caution or Youth Justice Conference, people with a victimisation event reported to the NSW police, people with both a contact as an offender and a victimisation event, and a cohort of people who experienced an episode in either remanded or sentenced custody. The proportions of people with disability who had contact with the criminal justice system are presented in Figure 6.4.2.

Figure 6.4.2. Proportion of people with disability in contact with the criminal justice system, by contact type



As shown in Figure 6.4.2, of those in receipt of a core disability service or support, 28% had criminal justice system contact as a victim, 16% as an offender, 10% as both a victim and an offender, and 5% had a custodial episode during the 10-year period[12].

The value of linking data to identify additional people with a disability is demonstrated below in Figure 6.4.3.

Figure 6.4.3. Proportion of people in contact with the criminal justice system with disability, by contact type

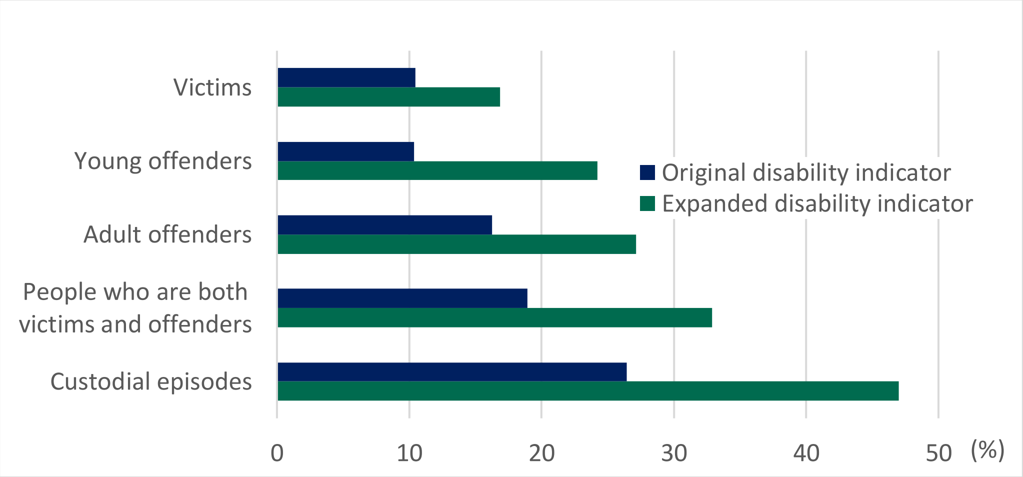


Figure 6.4.3 shows the proportion of people with different types of criminal justice system contact who were identified as having a disability based on their contact with one of the three

---

[12] Cohort groups are not mutually exclusive, people from one cohort group may appear in another.

core disability services or supports (original disability indicator). Using a broader definition of disability (expanded disability indicator), the proportion within each cohort who are identified as having a disability increases significantly. For example, just over 10% of young offenders were in receipt of a disability-specific service or support, but more than 20% were identified as potentially having a disability when additional data sources were considered.

The increase in the number of people with disability identified using the expanded indicator was largely due to an increase in the number of people identified as having a psychosocial disability. This is demonstrated in Figures 6.4.4 and 6.4.5 for the adult offending cohort and the custodial cohort, respectively.

Figure 6.4.4. Percentage of adult offending cohort with disability, by disability type

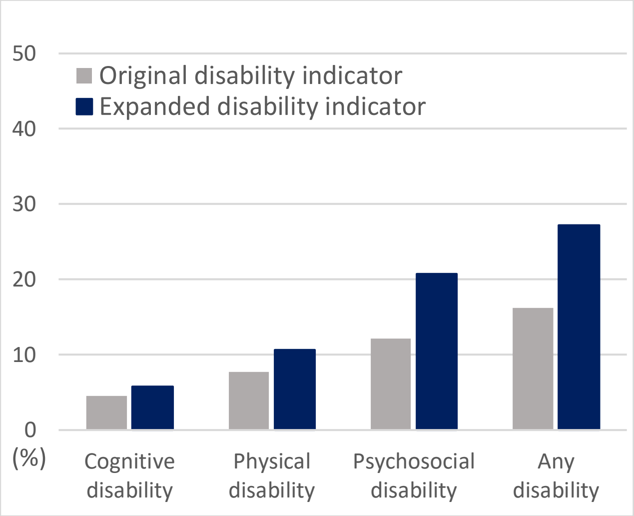
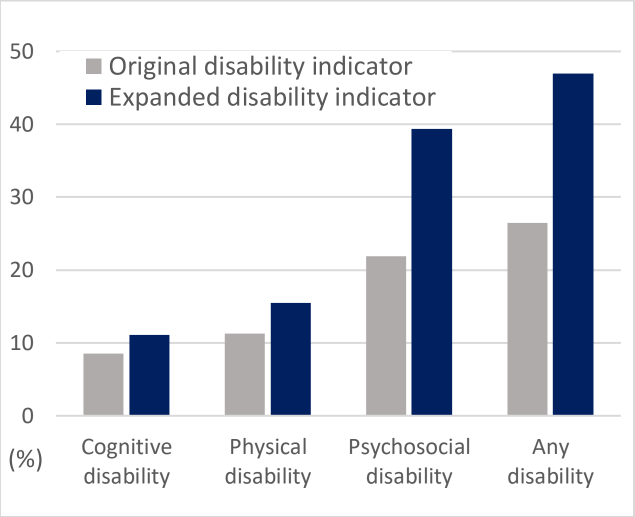


Figure 6.4.5. Percentage of custodial cohort with disability, by disability type



People with a disability may not be in receipt of a disability-specific service for a wide array of reasons, including that they may not meet eligibility criteria, may not be aware of their disability, or may not want to accept the service. These reasons can be particularly impactful for people in contact with the justice system as a victim or offender.

It is important to note that further refinements are needed to better identify people with disability, and these are detailed below. Results are interpreted in light of these caveats.

6.4.2 Data gaps and areas for data improvement

The methodology to identify people with disability could be improved in the short-term, both by enhancing the datasets currently available and by accessing information from additional data sources. For example, the inclusion of detailed descriptions of diagnoses from DOMINO and more complete hospitalisation data would enable more precise identification of people with disability. Further, sources of information for young offenders were limited in this test case. The inclusion of early childhood and education data and health service contacts, particularly prior to the age of 10 years, will likely be of value in the future. Due to the age criteria of disability-specific supports and services, it was also not possible to examine outcomes for older people with disability in this test case. Including data from a longer time period in the enduring asset would help to overcome this limitation.

This test case highlighted data development to better identify people with disability in contact with the justice system summarised below.

Table 6.4.3. Data development activities to improve identification of people with disability in contact with the justice system

Short-term	Inclusion of additional data which exist and are of moderate - good quality.	• Granular diagnostic DOMINO data • NDIS application date • Hospital data detailing, exact admission & separation dates, ICD codes • More complete data on area of residence • Education data • Additional categorisation of disability type from Health data sources • Further work to extract information from housing data • Community Mental Health services • Key medications from Pharmaceutical Benefits Scheme
Medium term	Data to be sought from additional sources. Data is of unknown quality (for linkage).	• Koori court • Court diversion programs • Non-government service and support system • Legal services (Legal Aid / ALS) • Young people in custody health survey • Inmate health survey • Wellbeing and quality of life measures
Long term	Information is missing or currently not well captured within administrative systems. It may also not be appropriate to capture some of this information in large administrative datasets.	• Cognitive and mental health assessments • Victimisation incidents not reported to police. • Disability outside of diagnosis: trauma; foetal alcohol spectrum disorder. • Complex support needs rather than diagnosis. • Cultural protective factors • Trauma and adverse childhood experiences • Family and community connections and experiences

Improvements to the indicator of severity and the temporal aspect of identification or acquisition require further development, beyond the time constraints of this test case. This test case identified a need for better quality metadata, more informative data dictionaries, and more specific details regarding exclusions from available datasets. Some data quality, identification and sequencing issues from this test case can be resolved with access to more granular data in individual datasets. Specifically, there is value in going back to state hospital collections to obtain the appropriate granularity of information.

Medium to longer term data improvements

A greater understanding of the supports needed for people with disability in contact with the justice system, including supports needed to prevent further contacts, will be provided by increasing the scope of data sources beyond government administrative data. Ultimately this will include the intersection of other risk factors (adverse childhood experiences, substance abuse, poverty and other vulnerabilities) and exploring family and community connections. Some of these are included in the table and will be included in the pilot deliverable: Data Development plan. Of note, many data gaps exist at the point of collection, information is not routinely collected, or there is no good existing administrative measure. These gaps particularly impact the population of this test case. Additional work is required on how best to combine the information produced from large administrative datasets with other information sources.

Indicators of disability in the NSW criminal justice system – A First Nations Perspective

First Nations people are over-represented in both disability and justice system populations. The team for this test case included an embedded Aboriginal Perspectives Expert Panel with broad ranging experience including social work, legal representation, advocacy, trauma research and healing, criminology and social justice. Importantly, this experience includes working directly with people in contact with the criminal justice system as victims, offenders, or both. The insights obtained from this group draw together insights from field research, interviews, and social work.

The analysts worked with the panel to plan the approach, interpret findings and identify data development areas particular to First Nations people. This model has proved effective in not only guiding this test case, but also informing the data development plan and enduring design.

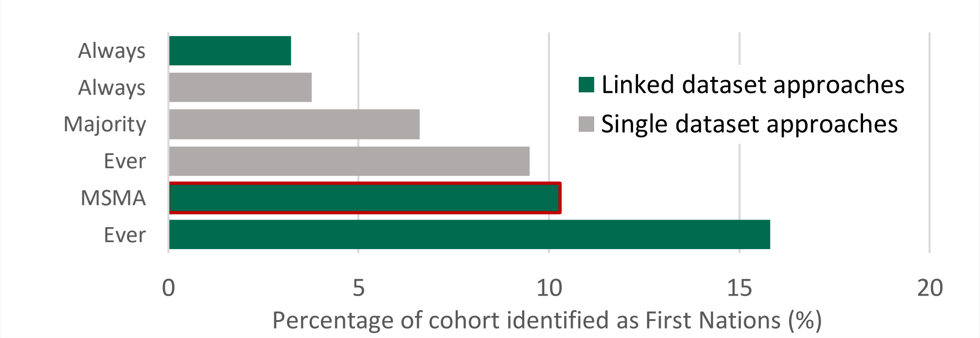
Identification of First Nations people in this test case

> ***"At our first workshop we recognised the issues of our mob not identifying,***
> ***but we did think it was important to be able to identify our mob in the data."***

The analytical team explored a number of methods for identifying First Nations people in the data and presented the impacts of these different identification methods to the panel.

- Most collections rely on self-identification. This can vary in one dataset overtime.
- Self-identification will be context dependent. e.g., people may be more willing to self-identify to a known health practitioner than police.
- In "Aboriginal identification: the way forward. An Aboriginal Peoples' Perspective (2015)", NSW Aboriginal Affairs surveyed 500 Aboriginal people about their self-identifying habits, finding that approximately 8% of people choose to never identify to administrative services.
- The Multi-Stage Median Algorithm (MSMA) approach - described in the "National Best Practice Guidelines for data linkage activities relating to Aboriginal and Torres Strait Islander People" (ABS & AIHW, 2012) - was identified by the analysis team as the preferred approach to be inclusive but not skewed by data entry error.
- This method looks within and across datasets to see if/ how often an individual was recorded as First Nations.

Figure 6.4.6. Proportion of adult offending cohort identified as First Nations under different approaches



Some data issues specific to First Nations people in this test case

An inherent problem with administrative data is that it is based on a person's appearance in a system not around a person. It is hard to identify people unless they present as requiring a support, or self-identify with a disability. Under-reporting (and under-supporting) of disability is particularly prevalent for First Nations people. This can be related to fear of greater surveillance, removal, and other consequences. There is no word for disability in First Nations languages, this is challenging not just for self-reporting, but as a concept.

A learning from this test case so far is to use a wide range of already scoped administrative sources across housing, health, social services, and education to identify whether someone identifies as a First Nations person, and discloses a disability.

Enduring design and data development from a First Nations perspective

The panel highlighted that diagnoses are interwoven with complex intergenerational trauma, for which we currently have no hard data in these administrative datasets. While mechanisms to better capture complex trauma are explored, the panel made the following recommendation:

> *"We can recommend that this work is understood, interpreted and conceptualised through a framework that includes complex intergenerational trauma, colonisation, and multiple needs, even if it is outside the actual datasets."*

In designing supports and services and evaluating outcomes the family and community needs to be factored in.

> *"You can't just work with the young fella in prison, you need to work with the whole family, the whole community."*

It is recognised that it will take time to develop the data and design the asset to adequately address these complexities, and developing a framework in which to conceptualise disability, support needs, and outcomes for First Nations people in the enduring asset and determining how this fits with governance, access, and use of the data is a key first step.

Data development would also focus on a wider range of data sources, particularly from community-controlled organisations where First Nations people are more likely to identify and to self-report a disability and accept a targeted service or support.

Next steps

- Exploration of some of the factors associated with an increased risk of people with disability having contact with the criminal justice system as victims and/or offenders.
- Exploration of the outcomes of charges involving people with disability as victims and/or offenders, including the extent to which diversionary options are used for offenders with mental health/psychosocial/cognitive impairments.
- An analysis of health services use of a custodial cohort leading up to and following a custodial episode.

6.5 Indicators of disability in the SA education system

The *Indicators of disability in the SA education system* (Education to Employment) test case, led by SA, will analyse pathways from education to employment for young adults with disability in South Australia. It aims to investigate pathways through school education, senior secondary education, tertiary education (VET and University) and subsequent employment outcomes and assess differences against young adults without disability.

A list of datasets included in this test case is provided in the Appendix.

Highlights and key accomplishments since June 2021

- Identified 185,479 students enrolled in year 10 in SA government schools between 2005 and 2019.
    - 10% of this student population are identified as having a disability (based on information captured in the education data);
    - 2.7% of the student population has joined the NDIS;
    - 5.9% of the students are identified as having a disability in the DS-NMDS data.
- Defined methodology to identify individuals with disability and severity of disability.
- Defined analytic categories for pathways analysis:
    - Single/multiple disabilities;
    - Medical condition vs disability vs no medical condition/disability;
    - Functional severity of disability.
- Defined outcomes along the education/work pathway.

Noting the recency of data access (July 2021), findings for this test case are preliminary only.

AT A GLANCE

Aim

To analyse pathways from education to employment for young adults with disability

What was found:

- Of 185,479 Year 10 students enrolled in SA public schools between 2005-2019:
    - 10% were identified as living with disability from the education data;
    - 5.9% were identified as living with disability from the DS-NMDS data;
    - 2.7% had joined the NDIS.

What's Next:

- Refinement of disability indicators to include measurement of 'functional' severity.
- Assessment and comparison of outcomes for individuals identified as living with disability vs those without

In-scope population for this test case

- South Australian individuals enrolled as year 10 (or equivalent) students in public schools during 2005–2019.

6.5.1 Identification of people with disability and methodology

The merging of several datasets recording disability can complement the information available in the SA enrolment data and allow the computation of various disability indices to be used in the analysis of education and work outcomes. Initial observations on the available data are outlined below.

Education data

- Available indicators of disability are education oriented and will need to be complemented with other information from other datasets. Indeed, in the data available for this test case disability is captured at a single point in time (year 10); this limits visibility of any change in severity among those identified as living with disability, and does not capture new disabilities that may appear after year 10, or that people chose not to disclose at the time of data capture.
    - Leveraging linked data will be instrumental in creating a more robust indicator of disability that can overcome these limitations.
    - This is particularly relevant in analyses of labour market outcomes.
- Three types of disability are most prevalent: Language and communication & speech, Intellectual, Autism/Asperger.
- The enrolment data captures information on severity of disability; further work is needed to fully understand how severity is defined[13].

---

[13] Furthermore, there seems to be a relative ‘inflation’ in the recorded ‘severity’ of disability for more recent year 10 census years: those who were in year 10 more recently are, on average, recorded with a more severe disability. We believe this ‘inflation’ is related to administrative considerations, more so than an actual worsening of students’ disability over time.

Further information is required from the Department of Education to understand the context in which data was collected (including historical changes in definition).

NDIS data

- A crude indicator of severity of disability can be obtained simply by identifying those students who have become NDIS participants or have made an access request.
- The NDIS data captures more detailed information on disabilities which can be reconciled with the ICD-10 coding framework. It also includes information on multiple disabilities and their types.
- 2.7% of the original student population has joined the NDIS. We note that there are individuals recorded as not having a disability in the SA School data that have been assessed as having a disability by the NDIS (Table 6.5.1).

Table 6.5.1. Disability status in the students file versus NDIS participation

	Disability status from the NDIS access met data					
	With no disability		With disability		Total	
Disability status from the SA School Enrolment database	No. of individuals	%	No. of individuals	%	No. of individuals	%
With no disability	166,180	99.6	719	0.4	166,899	100.0
With disability	14,366	77.3	4,214	22.7	18,580	100.0
Total	**180,546**	**97.3**	**4,933**	**2.7**	**185,479**	**100.0**

The NDIS flag is a crude indicator because NDIS participation is based on a range of eligibility criteria, not all having to do with limiting education/work outcomes. Moreover, earlier student cohorts are under-represented in the NDIS data.

Disability Services National Minimum Dataset (DS-NMDS)

- Preliminary investigation suggests that DS-NMDS data are the most promising source of information on disability in this test case. The data have two valuable dimensions:
  - Comprehensive classification of the type of disability (Table 6.5.2)
  - Information on the severity of a disability in terms of its impact on whether an individual needs help or supervision over nine domains of life (self-care; mobility; communication; interpersonal interactions and relationships; learning, applying knowledge and general tasks and demands; education; community (civic) and economic life; domestic life; working).

- Both discrete and continuous disability indicators can be constructed from the aggregation of individuals' answers to the nine domains. Other indicators can be constructed focusing the attention to a subset of domains such as help needed relating to education and work.
- The DS-NMDS fulfils many of the data requirements imposed by a comparative analysis of educational and work outcomes between young people with disability and other people.
- The data contains detailed information on disabilities which is updated every year for students accessing relevant supports. 5.9% of the students are identified as having a disability in the DS-NMDS data.

For each domain and individual with a disability we have a score indicating the severity of the functional impact of the disability (from 'unable to do or always needs help or supervision' to 'doesn't' need help or supervision and does not use aids or equipment'). Analysis indicates that there is a high correlation in the level of disability across the nine domains. **We conclude from this that a simple summary score equal to the average of the scores from the nine domains will be a robust measure of the severity of a disability. This summary measure can be used as a continuous variable or used to define categories of severity.**

Table 6.5.2. Distribution of primary disabilities in the DS-NMDS data

	No. of individuals	%
Intellectual	3,145	**29%**
Psychosocial	2,277	**21%**
Autism	1,704	**16%**
Specific learning/ADD	1,684	**15%**
Physical	1,032	**9%**
Neurological	228	**2%**
Not stated	209	**2%**
Acquired brain injury	200	**2%**
Vision	200	**2%**
Hearing	157	**1%**
Other disability (inc. deaf/blind, speech, developmental delay)*	92	**1%**
Total	**10,928**	**100%**

**: we grouped these disability types together because of possible small cells issues*

Other data sources

- Additional data linked for this test case include the DOMINO (Data Over Multiple INdividual Occurrences), Higher Education Information Management System (HEIMS) records, the Australian Apprenticeships Incentives Program (AAIP) data and the Total VET Activity (TVA) data
- The HEIMS, the AAIP and the TVA data provide information on self-reported medical conditions/disabilities. However, information is only available for students who have enrolled in university or vocational education and training (VET), or obtained an apprenticeship.
- Similar information on medical conditions/disabilities is also captured in DOMINO.
- Together, this data can be used to extract information on students who have a long-term medical condition (which is usually not recorded as a disability) and used to refine the disability categories used for analysis.

6.5.2 Data gaps and areas for data improvement

General issues

- No metadata (data dictionary) was available for data provided by the Department of Education; we are seeking input from the Department to further understand the data captured in each variable and to resolve data quality issues. Technical issues with the analytic environment have also delayed progression of the work.
- It is essential to use the underlying literature to justify the selection of data, how to code the information, and which models to employ.
- An important methodological consideration with the NDDA is the fact that the data has not been collected for research purpose. The administrative data have been collected by Departments and organisations for their own business and reporting purposes.
- These business rules may have some consequences on the types of analyses that can be achieved. Also, they have a direct bearing on the definition of disability. Some of the information on disability will be self-reported. Some will be recorded according to a range of pre-determined, usually policy or program related, criteria which apply to an individual at some point in time, notably to assess one’s eligibility for supports or payments. Another important consideration is that the definition of disability needs to be available for the whole population. This limits the use of information obtained from output data (participation in higher education, for example) that is not available for the whole population.
- The literature points to significant differences in terms of outcomes depending on whether someone has one or multiple disabilities. It would be useful if the NDDA allowed

researchers to distinguish between single and multiple disabilities. This information seems to be contained in the original population file, which is promising. Another important aspect of disability is that it is not static, especially for those who have multiple disabilities. What is recorded as the primary disability in the data for an individual may not actually be the one that exerts the most impact on educational or labour market outcomes. The primary disability may also change over time. Therefore, it would be useful for the NDDA to allow for updates to the information on disability over time.

- The age of onset of disability is an important determinant of educational and labour market outcomes. Whether someone was born with a disability has an important impact on outcomes. We are yet to find pertinent information in this test case relating to this aspect of disability.
- The literature also highlights the importance of chronic medical conditions (usually not classified as disabilities) in determining outcomes. Such conditions impact students' ability to attend class and may affect their ability to focus in class. These impact on educational and later outcomes. Presumably, the coding of disability in the original dataset being education oriented, the disability information should include both chronic conditions and disabilities. Cross tabulating that information with the more detailed information on disability from other datasets may allow one to identify those people who have a chronic condition in the original dataset.
- The severity of a disability is also key. In this regard, the NDIA data provides an opportunity to distinguish, in the original population, those individuals who experience severe and permanent disabilities. The DS-NMDS data is also promising in that it records the level of help and/or supervision a person with disability requires in nine life areas.
- Note that the original dataset tells us whether each individual has a disability (education oriented in the Department of Education data) reported in the census year. Looking at outcomes in later years, it will be useful to try and find updated information about disability throughout the years. At this early stage of the project, we focused on the datasets where some information about disability could be found. We have not merged outcomes yet.
- As the NDIS was rolled out, we notice that the number of people from more recent year 10 census years present in the DS-NMDS data decreases. These younger people who no longer appear in the DS-NMDS data seem to appear, instead, in the NDIA data as NDIS participants. We will need, somehow, to account for this when using the DS-NMDS data for constructing indicators of disability.

Data quality issues

- **Inconsistencies:** There are issues in the data about multiple records available for a number of students (4,254 students have more than one record). The census year 2012 has overwhelmingly more cases of multiple records than the other years. Some of the multiple records seem to provide contradicting information on items that should not be time variant. The team needs information on these issues from the Department of Education in order to determine whether they may affect the reliability of the data.
- **Time coverage of the datasets:** datasets like the TVA and, more importantly the DS-NMDS, do not cover the full 2005 to 2019 time range. The latter starts in 2011. While it records disability information for people of all cohorts including the older ones (2005 and later), over time changes occur as we can see in the period covered by the data. The issue is that we can observe these changes through the period covered by the data but not for the period 2005 to 2011. It would be useful for all datasets (except the NDIA, which came into existence in 2013) to cover the whole period 2005 to 2019 for the SA test case.
- **Population of interest:** The data covers only students enrolled in SA Government schools. It does not include students enrolled in Catholic and Independent schools. The issue with focusing on Government school students is that the analysis of outcomes only applies to this population. It cannot be generalised to the broader population of students without making some strong assumptions about the generalisability of the results. It is very likely that the students in non-Government schools (including those with a disability) are a self-selected group. First, on average, their parents belong to significantly higher socioeconomic groups (parents' socioeconomics is a strong determinant of educational outcomes). Second, the individual characteristics of students with disability enrolled in non-government schools are likely to be significantly different from those enrolled in government schools, including the characteristics (type, severity, etc.) of one's disability. We would recommend that information on cohorts of students enrolled in non-government schools be added to the data.
- **Additional information on disability:** to our knowledge additional information on disability is available from data collections. For example, the National VET Provider Collection contains information on disability type (see Karmel and Nguyen 2008), which would potentially be an improvement of the self-reported flag we currently have in the TVA data.
- **Incomplete Census data:** it appears that the Census data provided is incomplete with regards to relevant information on disability. Indeed, we have not been able to find the data relating to Question 23 of the 2016 Census which asks for the reason(s) why someone needs assistance or supervision.

Next steps

- To conduct a preliminary analysis of outcomes using the disability definition of the Department of Education (education oriented), simply distinguishing students with disability from those without.
- Refine the analysis using (recoded) indicators of severity (levels of disability) used by the Department of Education. At this stage, we would distinguish, among students with disability between those with a so-called 'additional' level (milder) and those with 'direct or above' recorded level (more severe).
- Combine the information from the DS-NMDS and NDIA to further improve the measurement of disability through a 'functional' severity measure. This will be done through both a continuous and discrete measure of functional severity computed from the nine domains of life in the DS-NMDS data.
- Improve the analysis of outcomes by refining categories of students further, namely by identifying students who are likely to have a long-term medical condition which is not recorded as a disability. This step will be achieved by combining HEIMS, TVA, AAIP and DOMINO data. The resulting analysis of outcomes will contrast individual pathways according to (i) single/multiple disabilities, (ii) medical condition vs disability vs. no medical/disability, and (iii) functional severity of disability.

7 Summary

A substantial amount of work has been undertaken across five policy test cases and laid the foundation for development of a suite of disability indicators that will form the cornerstone of an enduring asset. Preliminary results have indicated that linked administrative data can be used to develop reliable indicators of disability for adults aged 25-64. Disability indicators based on the linked data aligned closely with Survey of Disability, Ageing and Carers (SDAC) estimates for adults aged 25-64.

More work will be required to ensure disability data is fit-for-purpose. Further work is required develop reliable indicators for older people, children and young people, and disability related to complex trauma; this will require linkage of additional data and investments in data harmonisation and development. In addition, under-reporting and under-supporting of disability is prevalent among First Nations Australians; work led by First Nations Australians and investments in data development are required to address this gap. The benefits of ongoing investments in data improvement will extend beyond the NDDA to support other national initiatives, including Australia's Disability Strategy, Closing the Gap, and addressing critical gaps identified by the Royal Commission into Violence, Abuse, Neglect and Exploitation of People with Disability.

A third key learning has been a need to invest in data infrastructure. Any future infrastructure will need to have the capability to manage very large volumes of data and computationally intensive analysis, to enable refinement and derivation of high-quality, reliable disability indicators.

Further insights will be presented in a third series of test case reports in December 2021.

APPENDIX

Source	Dataset	Test cases 1	 2	 3	 4	 5
Commonwealth	Medicare Consumer Directory (MCD)	🟩	🟩	🟩	🟩	🟩
	National Disability Insurance Scheme (NDIS)	🟩	🟩	🟩	🟩	🟩
	Medicare Benefits Scheme (MBS)	🟩	🟩	🟩	🟩	
	Data Over Multiple Individual Occurrences (DOMINO)	🟩	🟩	🟩	🟩	🟩
	Child Care Subsidy (CCS)/Child Care Benefit (CCB)			🟩		
	Australian Early Development Census (AEDC)			🟩		
	Pharmaceutical Benefits Scheme (PBS)	🟩	🟩		🟩	
MADIP	Personal Income Tax (PIT) – ATO					🟩
	Higher Education Information Management System (HEIMS)					🟩
	Vocational Education and Training (VET)					🟩
	Survey of Disability, Ageing and Carers (SDAC)	🟩				
	Census	🟩				
National (AIHW)	Disability Services National Minimum Dataset (DS-NMDS) and Commonwealth State/Territory Disability Agreement National Minimum Data Set (CSTDA NMDS)	🟩	🟩	🟩	🟩	🟩
	National Death Index (NDI)	🟩	🟩	🟩	🟩	🟩
	Specialist Homelessness Services Collection (SHSC)	🟩			🟩	
	National non-admitted patient emergency department care database (NNAPEDCD)				🟩	
	National Hospital Morbidity Database (NHMD)				🟩	
	Public Housing and State Owned and Managed Indigenous Housing (PH & SOMIH) Data Set	🟩				
SA	SA School Enrolment Census					🟩
	SA Certificate of Education					🟩
VIC	Victorian Linkage Map (VLM 1906)		🟩			
	Victorian Integrated Data Resource (IDR), derivations and tabulations		🟩			
	Clinical public mental health services (CMI/ODS)		🟩			
	Mental health Community Services (MHCSS)		🟩			
	Victorian Admitted Episode Database (VAED)		🟩			
	Victorian Emergency Minimum Dataset (VEMD)		🟩			
	Victorian Integrated Non-Admitted Health (VINAH)		🟩			
	Community Health Minimum Dataset (CHMDS)		🟩			
	Home and Community Care (HACC)		🟩			

Test cases:
Test Case 1: Outcomes data (Cth DSS)
Test Case 2: Mental Health and Psychosocial (VIC)
Test Case 3: Early Childhood (NSW)
Test Case 4: Justice (NSW, Cth DSS)
Test Case 5: Education to Employment (SA)

		Test cases				
Source	Dataset	1	2	3	4	5
NSW	NSW-AIHW Linkage Map			🟩		
	NSW Education disability			🟩		
	NSW school enrolment			🟩		
	NSW community preschool census data			🟩		
	NSW Government Preschools and Early Intervention Census			🟩		
	NSW Preschool Disability Support Program (PDSP) – administered by Northcott			🟩		
	NSW Disability and Inclusion Program (DIP) - Higher Learning Support Needs (HLSN)			🟩		
	NSW Best Start			🟩		
	Nationally Consistent Collection of Data on School Students with Disability (NCCD)			🟩		
	Suspension data			🟩		
	Attendance data			🟩		
	NAPLAN data			🟩		
	NSW Birth Registration Data Collection			🟩		
	NSW Deaths Registration Data Collection			🟩		
	NSW Perinatal Data Collection (PDC)			🟩		
	NSW Emergency Department Data Collection (EDDC)			🟩		
	NSW Admitted Patient Data Collection (APDC)			🟩		
	NSW ChildStory			🟩		
	NSW Re-offending Database (ROD)				🟩	
	NSW Police Victims' records				🟩	
	NSW Social housing data				🟩	
	NSW Child Protection				🟩	
	NSW Out of home care (OOHC)				🟩	