Australian Government
Australian Institute of
Health and Welfare

# National Health Data Hub
# Data User Guide

Version 1.0
<September 2025>

# Table of Contents

## List of Abbreviations and Acronyms

| **Acronym/Abbreviation** | **Full Form** |
|---|---|
| ABS | Australian Bureau of Statistics |
| ACAP | Aged Care Assessment Program |
| ACFI | Aged Care Funding Instrument |
| ACMPS | Aged Care Management Payment Systems |
| AIHW | Australian Institute of Health and Welfare |
| AIR | Australian Immunisation Register |
| ACMPS | Aged Care Management Payment System |
| AN-ACC | Australian National Aged Care Classification |
| ANZICS | Australian and New Zealand Intensive Care Society |
| APD | Adult Patient Database |
| APPs | Australian Privacy Principles |
| CACP | Community Aged Care Program |
| CHSP | Commonwealth Home Support Programme |
| CORE | Centre for Outcome and Resource Evaluation |
| Cwth | Commonwealth |
| DAA | Data Access Agreement |
| DEX | Data Exchange |
| DoB | Date of Birth |
| DVA | Department of Veterans' Affairs |
| ECA | Early Childhood Approach |
| EACH | Extended Aged Care at Home |
| EACH-D | Extended Aged Care at Home Dementia |
| EDW | Enterprise Data Warehouse |
| ECEI | Early Childhood Early Intervention |
| EMS | Enhanced Medicare Spine |
| FY | Financial Year |
| GACDU | GEN and Aged Care Data Unit |
| GPMS | Government Provider Management System |
| HACC | Home and Community Care |
| HCP | Home Care Packages |
| HPV | Human Papillomavirus |
| ICU | Intensive Care Unit |
| ICD-10-AM | International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, Australian Modification |
| KBL | Key-Based Linkage |
| LHN | Local Hospital Network |
| MBS | Medicare Benefits Schedule / Medicare Benefits Scheme |
| METEOR | Metadata Online Registry |
| NCIS | National Coroners Information System |
| NACDC | National Aged Care Data Clearing House |
| NACDA | National Aged Care Data Asset |
| NAPS | National Approved Provider System |
| NDI | National Death Index |
| NDIA | National Disability Insurance Agency |
| NDIS | National Disability Insurance Scheme |
| NHDH | National Health Data Hub |
| NHDISC | National Health Data and Information Standards Committee |
| NHIA | National Health Information Agreement |
| NIHSI | National Integrated Health Services Information |

| | |
|---|---|
| NMDS | National Minimum Data Set |
| NBEDS | National Best Endeavours Data Set |
| NMLK | National Master Linkage Key |
| NNAP(el)D | National Non-admitted Patient Care Episode Level Database |
| NNAPEDCD | National Non-Admitted Patient Emergency Department Care Database |
| NPHED | National Public Hospital Establishment Database |
| HREC | Human Research Ethics Committee |
| NSAF | National Screening and Assessment Form |
| PBS | Pharmaceutical Benefits Scheme |
| PIC | Public Interest Certificate |
| PICR | Paediatric Intensive Care Registry |
| PIN | Personal Identification Number |
| PPN | Person-Project Number |
| PRL | Probabilistic Linkage |
| RAC | Residential Aged Care |
| RBDM | Registries of Births, Deaths and Marriages |
| RCS | Residential Classification Scale |
| RHCA | Reciprocal Health Care Agreement |
| RON | Research Only Network |
| RPBS | Repatriation Pharmaceutical Benefits Scheme |
| SAE | Secure Access Environment |
| SDC | Statistical Disclosure Control |
| SEAD | Secure Environment for Analysing Data |
| SLK-581 | Statistical Linkage Key 581 |
| SPARC | System for the Payment of Aged Residential Care |
| STRC | Short-Term Restorative Care |
| TCP | Transition Care Programme |

*The NHDH Data User Guide is a co-investment partnership with the Australian Research Data Commons (ARDC) (DOI: 10.3565/JA9K-4F12). The ARDC is enabled by the Australian Government's National Collaborative Research Infrastructure Strategy (NCRIS).*

# 1. Introduction

## 1.1 Purpose

The National Health Data Hub (NHDH) Data User Guide is a dynamic reference document designed to support both government and non-government users in understanding and using the NHDH. It provides essential, high-level information about the creation, structure, and ongoing development of the NHDH. As new data modules are added or linkage methodologies evolve, the guide will be updated periodically by the NHDH team to ensure accuracy and relevance.

## 1.2 Scope

This user guide introduces the NHDH to users seeking to access and use its data. It outlines the scope and coverage of the NHDH, detailing the core data modules, how data sets are collected, processed, and linked, and the methodologies that underpin these processes.

It also includes references to data quality statements, data item lists, technical notes and other relevant resources to support users in understanding and using the data effectively.

## 1.3 Key contacts

For information about the NHDH that is not covered in this user guide, please contact the NHDH team via NHDH@aihw.gov.au.

## 1.4 Background

The NHDH (previously known as the National Integrated Health Services Information (NIHSI)) is a major national data linkage system for health and welfare research and analysis. It comprises de-identified and longitudinal health and welfare data from state and territory, Commonwealth and non-government data sources.

The NHDH is unique among similar linkage assets as it integrates hospital data on admitted patient care services, emergency department services and outpatient services for most states and territories (except for Western Australia and the Northern Territory). The NHDH draws on data from nationally compiled data sets supplied by jurisdictions which are curated by the AIHW, making harmonised and nationally standardised linked health and welfare data available to analysts and researchers. It also includes various Commonwealth and non-government data sets relating to pharmaceuticals, health utilisation, aged care, deaths, disability, intensive care and immunisation information.

The NHDH facilitates person-focussed and longitudinal studies to support contemporary medical research, inform health, welfare, disability and aged care services planning and policy development by government and non-government organisations, and monitoring of service delivery.

# 2. NHDH Linkage Methodologies

## 2.1 Linkage of NHDH Data sets to the Enhanced Medicare Spine (EMS)

Data sets are individually linked to the EMS at the person level to create the NHDH. The EMS serves as the foundation of the AIHW linkage system for integrating health and welfare related data sets into the NHDH.

The EMS is created by linking personal identifiers from the Medicare Consumer Directory (MCD) with supplementary information from the National Death Index (NDI) and the Australian Immunisation Register (AIR). These linkages help update and enrich MCD records, ensuring that EMS provides a more complete and accurate foundation for health data linkage and analysis across the NHDH. The EMS plays a critical role in data linkage accuracy and interoperability between the Australian Institute of Health and Welfare (AIHW) and the Australian Bureau of Statistics (ABS) for including data in the PLIDA (where approval is granted to do so). While the EMS incorporates additional data from AIR and NDI no additional people are included from these data sources. The MCD remains the core reference file, as the EMS only includes individuals eligible for Medicare.

The data available for linkage on the EMS include full name, dates of birth and death, sex, contact address (which may be a post office box), postcode, and date of last use of Medicare Benefits Schedule (MBS) or Pharmaceutical Benefits Scheme (PBS).

EMS personal structure is defined by Services Australia. It is primarily based on Medicare Personal Identification Number (PIN) but takes into consideration the small number of cases where multiple PINs have been assigned to the same individuals. For example, an individual has multiple MCD records due to administrative duplication. The Australian Immunisation Register shows consistent vaccination history across both records, helping confirm they belong to the same person. This supports resolving duplicate PINs and improving person-level accuracy in the creation of the EMS.

The EMS covers most of the Australian population, as demonstrated by the high linkage rate achieved in linkages to national data sets. Some residents of Australia are known to not be covered, as they are not eligible to, or have not chosen to enrol in Medicare. These may include:

- Temporary migrants, including international students and people on temporary work visas and their families.
- Young children not enrolled at birth.
- Non-residents, such as international visitors.

It should also be noted that, as with all longitudinal administrative data, the coverage is affected by policy changes over time.

Linking to a population spine as opposed to separate data sets improves linkage quality and efficiency. The assumption that almost everyone in Australia is present in the EMS underlies the confidence placed in linkage results derived in this manner.

The available information varies between data sets. Data are linked to the EMS using three distinct processes—probabilistic name-based linkage or multi-step key-based linkage or deterministic merge—depending on the data available for linkage. These are described below.

### 2.1.1 Probabilistic Name-Based Linkage

Name-Based Linkage (NBL) is conducted using probabilistic linkage (PRL) techniques. This process involves generating potential record pairs by matching entries from two data sets based on common variables, such as personal names and other demographic attributes available in both sources.

Probabilistic linkage is essential when data sets lack a single unique identifier or when linkage variables are incomplete or inconsistent. It estimates the likelihood that two records refer to the same individual by comparing the similarity of their attributes. This method depends on combining several variables—such as first name, date of birth, and residential address that are sufficiently distinctive to enable matching in both data sets. Fields that can change over time, like postcode alone, are generally inadequate for reliable linkage. In many cases, it offers greater accuracy than deterministic methods by accommodating variability and uncertainty in the data.

### 2.1.2 Multi-Step Key-Based Linkage Approach

In situations where full names are not collected or not available to be shared, data sets typically include a common identifier at the level of the individual, such as the Statistical Linkage Key (SLK-581)—. To enable linkage under these conditions, a multi-step key-based linkage (KBL) method is applied. This approach involves the following:

- Primary linkage variables:
    - SLK-581: A standardised key derived from selected personal attributes.
    - Postcode: Used in conjunction with SLK-581 to enhance linkage accuracy.
- Use case:
    - Used when full name data are unavailable or incomplete.
    - Essential for maintaining privacy while still enabling low sensitivity and high specificity person-level linkage.
- Advantages:
    - Supports linkage across data sets lacking direct identifiers.
    - Enables data use while retaining confidentiality.

The KBL methodology ensures consistency, reliability, and comparability in data integration processes.

### 2.1.3 Deterministic Merge Approach

A deterministic merge is a method of combining data sets where records are linked only when there is an exact match on a key identifier - in this case, a Person Identifier Number (PIN). For example, if the PIN in MBS or PBS is the same as the PIN in the EMS, the records are automatically joined. This differs from other methods (like probabilistic linkage) that rely on similarity or probability across multiple fields (e.g., names, date of birth). A deterministic merge removes any uncertainty because it uses an exact match on the PIN. This method is used to link MBS, PBS and AIR records to the EMS, ensuring records are accurately connected to the correct individual's EMS record.

## 2.2 NHDH Data Design

Common data variables across data modules in the NHDH are treated to meet data custodian conditions and to help mitigate against the risk of identification or spontaneous recognition of individuals, sensitive sub-populations, and organisations in the national linked data system. For the NHDH, a common variable is defined as a variable that is present in more than one data module (e.g. personal characteristic, geography data) or a date variable that interfaces with another date variable (e.g. separation date in hospitals data may be an entry date in aged care data or a date of death in NDI data).

The Data Design and Expansion Team works closely with the data custodian on establishing the data item list for the NHDH and develop data specifications to transform the data in line with the requirements of the NHDH. The design of the NHDH is based on default and by-exception modules, where more sensitive data items are placed in the by-exception data modules and require additional justification in the project proposal to access this data.
Common variable treatments include, but are not limited to:

- Dates – Month and year are included in the default module, while full dates are included in the by-exception module.
- Postcodes – no postcodes are allowed in the NHDH, however Remoteness, SEIFA, SA2, SA3 and SA4 are derived from the postcode and included in the NHDH default data module.
- Country of birth and Language - 4-digit County of Birth for Australia and New Zealand are provided in the NHDH on the default data module and all other codes reflect 1-digit classification; and 4-digit Country of Birth for all records are provided in the NHDH on a separate by-exception data module.
- IDs – All IDs supplied by the data provider are either removed or scrambled.
- Personal information – all personal information is removed.
- Free text fields – all free text fields are removed.

## 2.3 Working with the NHDH data

In working with data from the NHDH, it is essential to understand how individuals and records are identified and linked across various data sets. This section defines the key identifiers used to manage and integrate data: PPN (Person-Project Number), ROW_ID, and SYNTHETIC_ID. These identifiers help ensure data consistency, support linkage across data sets, and enable analysis even when full linkage to the EMS is not possible.

| **Item** | **Definition** |
|---|---|
| PPN (Person-Project Number) | PPN is a unique number given to a person identified in Medicare Consumer Directory and based on the personal structure provided to AIHW by Services Australia. PPN is only created for individuals enrolled with Medicare. A PPN is unique for each person across linked data sets in the NHDH. A record will only be allocated a PPN if the record has linked to the EMS. A record without a PPN indicates the record has not linked to the EMS and is referred to as an unlinked record/episode. |
| SYNTHETIC_ID | A synthetic ID is assigned to a record to assist with understanding which records relate to the same individual across tables within a topic (e.g. aged care, Child protection) where data tables relate to each other. The synthetic ID is particularly important when the record has not linked to the EMS and hasn't been assigned a PPN.<br>The synthetic ID usually replaces the original ID supplied to the AIHW by data providers.<br>A synthetic ID will only relate to the same synthetic ID within a topic of data for national data (e.g. aged care HACC data) or within the same state and topic of data |

| | |
|---|---|
| | for state-based data (e.g. NSW Child Protection data).<br>A synthetic ID will be allocated to records with or without a PPN.<br>A synthetic ID cannot always be used to generate unique counts of people as they are not linked to the spine– meaning 2 synthetic IDs might relate to the same person. |
| ROW_ID | A unique row id is assigned to all episode-based data tables and is used to distinguish records that belong to the same PPN or Synthetic ID. It supports the merging of data tables that have been split in the NHDH (e.g. default and by exception data tables). Example: When combining hospital admission records across years, use ROW_ID to keep admission distinct even for the same person. |

Understanding these IDs is crucial for:

- Interpreting how records relate to individuals across different data sets and topics.
- Merging and managing episode-based data tables.
- Identifying when data are partially linked or unlinked, which affects data completeness and analysis.
- Practical rule of thumb: Use the PPN whenever it is present to build cross-program longitudinal views at the level of the individual; fall back on the synthetic ID when the PPN is missing to preserve within topic continuity. Together they provide full coverage, synthetic IDs protect relational integrity for unlinked cases, and PPNs knit linked cases into the broader NHDH linkage system.

These definitions apply throughout the documentation, especially in sections that describe data linkage, table structures, and data quality.

# 3. NHDH Content Data

## 3.1 NHDH Data Modules

The NHDH includes de-identified content data from the following data sources:

- Medicare Benefits Schedule (MBS) data
- Pharmaceutical Benefits Scheme / Repatriation Pharmaceutical Benefit Scheme (PBS/RPBS) data
- Hospitals data
    - Admitted patient care data
    - Non-admitted patient care data
    - Emergency data

- Australian Immunisation Register (AIR) data
- National Death Index (NDI) data
- National Aged Care data
- National Disability Insurance Scheme (NDIS) data
- Australian and New Zealand Intensive Care Society (ANZICS) data
- Patients Demography data

The following sub-sections contain an overview of each data module in the NHDH. These descriptions are updated by the Data Custodians with assistance from the Unit with primary responsibility for NHDH, and the NHDH Data Custodian.

For each module, the following information is provided:

- How NHDH content data are linked to the EMS
- How NHDH content data modules are updated
- Sources of NHDH content data
- When NHDH content data modules are updated
- What naming conventions are used for NHDH data modules
- Other methodological considerations or data information (where applicable)

### 3.1.1 Patients Demography Data

*Introduction*

The Patients Demography data module in the NHDH is a central data set derived by AIHW, providing core demographic information for all individuals who have a record linked to the EMS. This module includes demographic data from all linked source data sets to create a comprehensive, de-identified person-level profile within the NHDH.

*Source*

The module is updated with every release, incorporating new individuals as data sets expand. Updates to specific demographic attributes occur periodically, with Date of Birth (DoB) and Sex refreshed based on (MCD- data in September releases and Indigenous identification updated using hospital records in December. Historical versions of the demography file are retained for reference.

It is important to note that the patient demography data module does not include records from quarantined data sets, such as Mental Health and Perinatal modules, which are project-specific and not universally accessible. Instead, demographic information for these projects must be sourced separately.

This module plays a critical role in structuring the NHDH by ensuring that demographic attributes are consistently derived and maintained across data sets while adhering to defined linkage and updated methodologies.

*Scope*

The Patients Demography data module in the NHDH is derived by the AIHW and includes all linked records from across all core data modules in the NHDH. It includes three key data items: DoB, Sex, and “Ever Identified as Indigenous”.

- DoB and Sex are derived from the Medicare Consumer Directory (MCD). May 2024 release used MBS and PBS data. From October 2024 releases, derived records from the MCD include everyone linked to MCD having at least one NHDH event in scope since 2010-11. From March 2025, the start date for MBS/PBS was moved back to 01-Jul-2002.
- The "Ever Identified as Indigenous" data item is derived from Hospitals data.
- All other content data such as MBS, PBS, NDI, AIR, Aged Care (September 2024), and ANZICS (March 2025) are used to define the scope of Person Project Numbers (PPNs).
- From May 2025, the NDI start date was moved back to 1 July 1997.
- From March 2025, the MBS and PBS start dates were moved back to 1 July 2002.

Importantly, the Patients Demography module only include individuals who have had at least one linked episode of care or interaction captured in the NHDH data modules. Records without a linked PPN—i.e., those not linked to any NHDH event—are excluded from this data set.

*How Demography data is updated*

- Regular Updates: A new version of the Patient Demography data set is created with each release, incorporating new individuals from added data sets or extended temporal scope of the existing data sets.
- Data Corrections: Updates may include corrections to date of birth or sex, especially if there are changes from MCD.
- Indigenous Status: The March 2025 release includes updates to the "Ever Indigenous" flag based on the latest hospital data.
- Retention of Versions: All previous versions are retained and remain accessible to researchers.
- Exclusions: The data set does not include linked records from quarantined data modules like NHDH +n projects and pilot projects, due to access restrictions and resource limitations.
- Researchers working with quarantined modules must source demographic details (e.g., DoB, Sex, Indigenous status) from those specific modules or derive them from other NHDH core data sets.

*When Patient Demography data is updated*

The Patient Demography module is updated with every quarterly data release – March, May, August, November, which occurs regularly throughout the year. Updates typically align with:

- MCD updates – Annually in the September quarter, with plans to move to 6-monthly updates.
- Hospital data updates – Every November.
- ABS Cause of Death updates – In the September and March releases.
- New core data additions – For example, historical MBS/PBS data added in March 2025.
- NDI death data expansions – Such as the May 2025 addition of deaths from 1997–2002.

Each of these events can trigger updates to the Patients Demography module, especially when new Person-Project Numbers (PPNs) are introduced.

*Naming convention of demography data module*

Each version of the PATIENTS_DEMOGRAPHY file is named with a suffix indicating the month and year of the release, following this format:

- PATIENTS_DEMOGRAPHY_MAY24
- PATIENTS_DEMOGRAPHY_OCT24
- PATIENTS_DEMOGRAPHY_DEC24
- PATIENTS_DEMOGRAPHY_MAR25
- PATIENTS_DEMOGRAPHY_MAY25

This convention helps clearly identify the release date of each version for tracking and reference purposes. More detailed information in the Patients Demography can be found from the NHDH data item list.

### 3.1.2 Medicare Benefits Schedule (MBS) Data Module

*Introduction*

Medicare is Australia's universal health insurance scheme. It guarantees all Australians (and some overseas visitors) access to a wide range of health and hospital services at low or no cost. Medicare commenced in 1984 to help Australians with the costs of their health care, regardless of where people live or their ability to pay. It has since grown to cover a broad range of health services. The services covered by Medicare are contained in the Medicare Benefits Schedule (MBS). Medicare benefits are payable only for services that are listed on the MBS. A medical service is a clinically relevant service if it is accepted by the medical profession as necessary for the appropriate treatment of the patient.

The extent to which the Medicare subsidy covers the cost of a Medicare service depends on factors such as whether the service is in hospital or outside of hospital (determining which Medicare benefit rate is applicable), and how much the health practitioner charges. Health practitioners have the option to either bulk bill or to charge a patient for the cost of the service. Bulk billing means a patient does not have to pay for the service. Instead, a health practitioner accepts the Medicare benefit as the full payment for the service. Under this arrangement, the patient assigns their entitlement to a benefit to the health practitioner, and the health practitioner cannot charge the patient any additional fee.

Alternatively, a health practitioner may charge the patient a fee for the service. In this instance, the patient either pays the full cost or the difference between what the charge is and what Medicare covers. If the patient pays the full cost, they can then make a claim for a Medicare benefit. If you are a public patient in hospital, Medicare covers all your medical expenses. However, if you are a private patient in a public hospital, Medicare only subsidises the doctor fees paid by the patient and the patient pays for hospital accommodation, theatre fees and doctors' fees.

*Scope*

MBS data only includes services for which an MBS benefit was paid. They do not include services:

- Provided by hospital doctors to public inpatients
- Provided under the Department of Veterans' Affairs National Treatment Account
- Covered by third party or workers' compensation
- Delivered to repatriation beneficiaries or defence personnel

- Delivered for insurance or employment purposes
- That are funded directly by other Australian Government programs (such as health screening services)
- Funded directly by State/Territory Government programs.

The Australian Institute of Health and Welfare (AIHW) has access to a view of the data set held by the Australian Government Department of Health, Disability and Ageing (the Department) for approved studies, including data linkage projects.

*Source*

MBS claims data are an administrative by-product of Services Australia's administration of Medicare. MBS services are available to all Australian residents who hold a current Medicare card and overseas visitors from countries with which Australia has a Reciprocal Health Care Agreement (RHCA). Medicare benefits are generally only payable for professional services that are clinically relevant, fulfil all item requirements for the relevant MBS item, are provided by an appropriate practitioner permitted to participate in Medicare, and are provided in Australia. In addition, the patient needs to be eligible for Medicare. If all requirements for a service are met, a claim may be submitted to Services Australia for a payment of the Medicare benefit.

Services Australia administers the payment of Medicare benefits and collects MBS claims data under the *Medicare Act 1973.* These data are then regularly provided to the Department. The Department is the data custodian of the MBS data. The AIHW accesses MBS data from the Department's Enterprise Data Warehouse.

*Using MBS data*

MBS data can be used to measure whether a medical service was provided, the date of the service, and the benefit paid by Medicare. These records are useful for understanding service utilisation and healthcare costs. However, MBS data do not capture the clinical reasons for the service, any referrals or follow-up arrangements, or the outcomes of tests and procedures. Additionally, services provided to public hospital inpatients or those not listed on the MBS, or where eligibility criteria are not met, are not included in the data set.

For MBS data, choosing the right codes requires checking the schedule regularly, as items are updated quarterly. Codes may not always serve as accurate proxies for diagnoses, and some services may be under-captured due to billing practices like MBS coning, where only the most expensive pathology tests are billed. State and territory differences also affect how services are classified and recorded. When selecting the right population, researchers must ensure that individuals meet the eligibility criteria for subsidised services, as some items are restricted to specific clinical groups. Leakage—where services are provided outside of subsidy conditions—can result in missing data. Defining the study period involves accounting for seasonal trends, with MBS claims typically peaking in winter and dipping during holidays. It's also important to determine when a treatment series begins or ends, which may not be straightforward. Researchers should define start and stop points based on the nature of the service and use consistent criteria, such as identifying the first claim or using gaps in service to indicate cessation.

Analysing MBS data requires a tailored approach depending on the specific service or condition being studied. There are no standard methods, so researchers must develop rules that make sense for their research question and clearly document all assumptions. MBS data sets are large and complex, often requiring significant time and computing resources to prepare. Analysts should never modify the raw data directly and should expect to rebuild the final data set multiple times. When analysing MBS data, it's important to consider factors such as seasonal variation in service use, changes in item listings or restrictions, and the potential for “leakage” where services are provided outside of subsidy conditions and thus not captured. Consulting clinicians and coders is essential to understand how services are delivered and recorded. Researchers should also validate their findings against external sources like MBS statistics and published studies and be cautious not to infer causation from correlation.

*How MBS content data are linked to the EMS*

The MBS data are linked to the EMS using Personal Identification Numbers (PINs). The MBS is administrative data set, and its records contain PINs. They match the PINs in Medicare Consumer Directory held by Services Australia. These shared PINs ensure interoperability between EMS and MBS throughout a deterministic merge as mentioned in section 2.1.3.

*How MBS content data modules are updated*

Keep identifiers separate to content from data custodians and holdings are merged using linkage maps supplied by the AIHW linkage team.

*When MBS data modules are updated*

Content data modules are updated annually.

*Data reference period and states/ territories coverage*

The MBS contains all states and territories data from 1 July 2002 to 30 June 2023 as of May 2025 NHDH release.

*Naming conventions for MBS data modules*

MBS data module has been split into a content table and a by-exception table module as follows:

- MBS_CONTENT
- MBS_CONTENT_BYEX

Detailed information on the variables included in these MBS data modules is available in the NHDH data item list.

*Other data information*

There are two supplementary tables associated with MBS data modules: the MBS Item Map and the MBS Registered Specialty Codes. The MBS Item Map supports analysis of MBS data over time by linking original items to their most recent comparable versions, ensuring consistent tracking and comparison despite item remapping. The MBS specialty code supports analysis by identifying the specialty or approval used to assess a claim, which may differ from the provider’s main specialty.

In extremely rare cases (fewer than 0.001% of rows), the ORIGINAL_ITEM code in the MBS_CONTENT table may not have a corresponding entry in the MBS_ITEM_MAP. These instances are due to administrative data errors.

Because these records may reflect lower data quality overall, we recommend using caution when including these records in your analysis. Importantly, these records do not represent actual Medicare benefit payments, so excluding them should have little to no impact on most analyses.

For example, in the MBS content file, there is a single record from 2005 where the variable ORIGINAL_ITEM has the value “PA”. This value does not appear in the MBS item map, and therefore cannot be mapped or meaningfully interpreted. The record also does not contain an associated benefit amount. As this affects only one record of 9.8 billion, the impact is negligible if this record was removed from analysis.

*Other Methodological considerations*

The following two links provide some methodological considerations when analysing MBS data for linkage projects.

- Medicare Benefits Scheme funded services: monthly data, Technical notes - Australian Institute of Health and Welfare
- Medicare bulk billing of GP attendances: monthly data, Technical notes - Australian Institute of Health and Welfare

### 3.1.3 Pharmaceutical Benefits Scheme (PBS)/ Repatriation Pharmaceutical Benefit Scheme (RPBS) Data Modules

*Introduction*

The Pharmaceutical Benefits Scheme (PBS) and the Repatriation Pharmaceutical Benefits Scheme (RPBS) are Australian Government Health programs that subsidise the cost of a wide range of medicines in Australia. The PBS began in 1948 and is available to current Medicare card holders as well as to overseas visitors from countries with Reciprocal Health Care Agreements with Australia. The RPBS was established in 1919 for returning Australian service people who had served in previous wars. Veterans, war widows and widowers, and their dependants are eligible for the RPBS if they have an assessed clinical need for the item and hold a Department of Veterans' Affairs (DVA) Gold, White or Orange card.

*Scope*

The PBS data collection contains information on prescription medicines that qualify for a benefit under the *National Health Act 1953* and for which a claim has been processed. The database comprises information about PBS scripts and payments, patients, prescribers and dispensing pharmacies. PBS statistics only include prescription medicines that are listed on the Schedule of Pharmaceutical Benefits (PBS Schedule) and have had a claim processed.

They do not include:

- Private prescriptions, off-label prescribing, over-the-counter medicines
- Medicines supplied to public hospital in-patients (except New South Wales and the Australian Capital Territory)
- PBS Opiate Dependence Treatment Program prior to 1 July 2023
- Any PBS medicines supplied under special arrangements that are not processed through PBS online
- Information on reasons for prescribing, prescribed dosage, frequency of administration and intended duration – information on how the patient uses the dispensed medicine once the patient has received the medicine.

The Australian Institute of Health and Welfare (AIHW) has access to a view of the data set held by the Australian Government Department of Health, Disability and Ageing (the Department) for approved studies, including data linkage projects.

*Source*

PBS medicines data are an administrative by-product of Services Australia's administration of Medicare. PBS subsidised medicines are available to all Australian residents who hold a current Medicare card and overseas visitors from countries with which Australia has a Reciprocal Health Care Agreement (RHCA). The PBS has been in existence since 1948 and is governed by the *National Health Act 1953* (Commonwealth). The Department is the data custodian of the PBS data.

The AIHW accesses PBS data from the Enterprise Data Warehouse (EDW) provided by the Department.

*Using the PBS/RPBS data*

The PBS/RPBS data set is a robust and cost-effective resource that captures a large portion of the Australian population, offering substantial statistical power for research. It provides detailed information on community-based prescription medicine dispensing, including demographics of patients and prescribers, financial aspects, dispensing environments, and health conditions where relevant. Since its national coverage began in 2002 (with under co-payment data from 2012), it has become a valuable tool for investigating the Quality Use of Medicines (QUM) principles. Researchers can explore appropriate medicine choices, safe and effective usage, and conduct pharmacoeconomic studies. The data set supports longitudinal analysis, enabling the study of prescribing trends over time and the use of medicines in rare diseases or specific sub-populations. Additionally, PBS/RPBS data can be linked with other health data sets such as MBS, AIR, and hospital records, enhancing its utility in integrated health research projects.

Despite its strengths, the PBS/RPBS data set has several limitations. It is an administrative data set not originally intended for research, which introduces potential biases and confounding factors. Researchers must be cautious and consider guidelines when designing studies. The data set contains sensitive information and is subject to strict privacy regulations under the National Health Act 1953. It requires specialised knowledge to interpret correctly and lacks certain critical data points, such as the date of death, which can affect analyses involving long-term patient outcomes. Furthermore, it does not include private prescriptions, over-the-counter medicines, non-PBS listed drugs, medicines supplied to public hospital inpatients, off-label prescribing, and drugs under special arrangements like the PBS Opiate Dependence Treatment Program or the Special Access Scheme. It also lacks information on orphan drug designations, reasons for prescribing, and detailed usage instructions such as dosage, frequency, and duration. These gaps can limit the scope of research and necessitate data linkage with other sources to fill in missing information.

PBS data allow researchers to measure whether a prescription was filled, the date it was written and dispensed, the name and form of the medicine, and the associated costs to both the government and the patient. This makes PBS data valuable for tracking medicine use and expenditure. However, PBS data do not indicate whether the medicine was actually taken, the intended dosage, the duration of use, or prescriptions that were written but never filled. Private prescriptions and the medical reasons for prescribing a drug are also not captured, limiting the ability to assess clinical outcomes or indications.

Analysing PBS data requires a flexible and thoughtful approach tailored to the specific research question. There are no standard methods, as each medicine, condition, and population may require different assumptions. Researchers should treat each prescription and disease as unique, developing rules that make sense for their study and clearly documenting all decisions and assumptions. Sensitivity analyses are essential to test how different assumptions affect results. PBS data sets are large and complex and require substantial computing power and careful data set preparation. Analysts should never modify the raw data directly and should expect to rebuild the final data set multiple times.

Key considerations include understanding beneficiary status, which affects data capture—especially for below co-payment medicines that were not recorded for general beneficiaries before April 2012. Seasonal patterns also influence PBS data, with claims peaking at the end of the calendar year due to the safety net effect and dropping in January. Policy changes, such as the 60-day dispensing rule introduced in 2023, can alter item codes and usage patterns, so it's important to track these changes. Analysts should consult clinicians, pharmacists, and consumers to understand prescribing practices and validate findings against external sources like PBS statistics, OECD data, and published studies. Finally, researchers must be cautious in interpreting statistical associations, remembering that correlation does not imply causation.

*How PBS/RPBS content data are linked to the EMS*

The PBS data are linked to the EMS using Personal Identification Numbers (PINs). The PBS is ministrative data set, and its records contains PINs. They match the PINs in Medicare Consumer Directory held by Services Australia. These shared PINs ensure interoperability between EMS and PBS throughout a deterministic merge as mentioned in section 2.1.3.

*How PBS/RPBS content data modules are updated*

Keep identifiers separate to content from data custodians and holdings are merged using linkage maps supplied by the AIHW linkage team.

*When PBS/RPBS data modules are updated*

Content data modules are updated annually.

*Data reference period and states/ territories coverage*

The PBS contains all states and territories data from 1 July 2002 to 30 June 2023 as of May 2025 NHDH release.

*Naming conventions for PBS/RPBS data modules*

PBS data module has been split into a content table and a by-exception table module as follows:

- PBS_CONTENT
- PBS_CONTENT_BYEX

Detailed information on the variables included in these PBS data modules is available in the NHDH data item list.

*Other data information*

There are two supplementary tables associated with PBS data modules: the PBS Item Map and the PBS Major Specialty Codes. The PBS Item Map supports analysis by linking PBS item numbers to their associated drug names, forms, strengths, and Anatomical Therapeutic Chemical (ATC) 5 classification codes, enabling consistent identification and classification of medicines over time. The PBS Major Specialty Code identifies the subspecialty of the authorised health professional who prescribed a PBS medicine, helping analysts understand prescribing patterns by specialty.

*Other Methodological considerations*

The following link provides some methodological considerations when analysing PBS data for linkage projects. Pharmaceutical Benefits Scheme prescriptions: monthly data, Technical notes - Australian Institute of Health and Welfare

*Further information*

More information about the PBS data can be found here: Medicines Overview - Australian Institute of Health and Welfare

### 3.1.4 Hospitals Data Modules

*Introduction*

The AIHW collects and manages hospital data from states and territories as part of the annual reporting purposes. Within the NHDH, hospital data are not collected separately but they are reused from the AIHW's existing hospital data collections. By reusing AIHW supplied hospital data, the NHDH supports broader data linkage and integration, enabling richer insights into patterns of health service across multiple data sets.

This approach ensures consistency with national reporting standards while reducing duplication of data provision by states and territories.

*Scope*

The NHDH brings together data on admitted patient care services (in public and some private hospitals), emergency department services and outpatient services in public hospitals for all states and territories except for Western Australia and the Northern Territory.

Coverage of private hospitals data within the NHDH is limited and depends on the availability of identifiers provided by states and territories to the AIHW for its data linkage activities. Only Vic, Qld and the ACT have private hospitals data included in the NHDH and only for selected reference years. For South Australia, NHDH only includes admitted patient care and emergency department services in public hospitals.

*Source*

Data are sourced from the following four national collections compiled by the AIHW:

- **The National Non-Admitted Patient Emergency Department Care Database (NNAPEDCD)**

  A compilation of episode-level records (including waiting times for care) for non-admitted patients registered for care in emergency departments in public hospitals.

- **The National Hospital Morbidity Database (NHMD)**

  A compilation of episode-level records from admitted patient morbidity data collection systems in Australian public and private hospitals.

- **The National Non-admitted Patient Care Episode Level Database (NNAP(EL)D)**

  This holds episode-level information on non-admitted patient care provided by public hospitals, Local Hospital Networks (LHNs) and selected other public hospital services managed by state and territory health authorities.

- **The National Public Hospital Establishment Database (NPHED)**

  This holds information on public hospital resources and includes information reported for public hospitals, LHNs and state and territory health authorities.

These data are sourced from state and territory Health Authorities according to the relevant NMDS/NBEDS requirements and jurisdictions are primarily responsible for the quality of the data they provide.

*Using hospitals data*

The NHDH hospitals data align well with the estimates of some published reports. For example, the NHDH data are aligned with AIHW published tables: Table 1.1 from the *Non-Admitted Patient Care* report, Table 2.2 from the *Emergency Department Care* report, and Table 2.2 from the *Admitted Patient Care* report. Note that in the published admitted care data, certain care types (e.g. 7.3, 9.0, 10.0) are excluded, whereas the NHDH data includes all care types—so these must be filtered out to ensure consistency and full data alignment with the published hospital report.

*How Hospitals content data are linked to the EMS*

The Hospitals content data modules are probabilistically linked to the EMS using names, dates of birth and addresses via probabilistic name-based Linkage as discussed in Section 2.1.1.

*How Hospitals content data modules are updated*

Identifiers and content from data custodians and holdings are merged using Linkage Maps supplied by the AIHW Linkage team.

*When Hospitals data modules are updated*

Content data modules are updated annually.

*Data reference period and states/ territories coverage*

| **Hospitals data collection** | **States/territories included** | **Reference period** |
|---|---|---|
| Admitted patient care data drawn from the National Hospital Morbidity Database (NHMD) – public hospitals (including chronic conditions) | NSW, Vic, Qld, Tas, ACT<br>SA<br>NSW, Vic, Qld, Tas, ACT (chronic conditions)<br>SA (chronic conditions) | 2010–11 to 2022–23<br>2010–11 to 2021–22<br>2015–16 to 2022–23<br>2015–16 to 2021–22 |
| Admitted patient care data drawn from the NHMD – private hospitals (including chronic conditions) | Qld*<br>Qld* (chronic conditions)<br>ACT**<br>ACT** (chronic conditions)<br>Vic**<br>Vic** (chronic conditions) | 2010–11 to 2022–23<br>2015–16 to 2022–23<br>2010–11 to 2018–19<br>2015–16 to 2018–19<br>2010–11 to 2016–17, 2021–22 to 2022-23<br>2015–16 to 2016–17, 2021–22 to 2022-23 |
| National Non-Admitted Patient Emergency Department Care Database (NNAPEDCD) | NSW, Vic, Qld,Tas, ACT<br>SA | 2010–11 to 2022–23<br>2010–11 to 2021–22 |

| | | |
|---|---|---|
| National Non-Admitted Patient episode-level Databases (NNAP(el)D) | Tas, ACT | 2013–14 to 2022–23 |
| | ACT | 2013-14 to 2021-22 |
| | Vic, Qld | 2016–17 to 2022–23 |
| | NSW | 2017–18 to 2023 |

*Naming conventions for Hospitals data modules*

The naming conventions are as per the agreed metadata in METEOR, the registry for Australian metadata standards for statistics and information in areas such as health, housing and homelessness, aged care, First Nations health, disability, children and families, and youth justice.

The health registration authority endorses national metadata under the National Health Information Agreement (NHIA). The endorsing body is the National Health Data and Information Standards Committee (NHDISC).

For example, the relevant metadata for the 2022-23 collections are found here:

- Non-admitted patient emergency department care NMDS 2022-23
- Admitted patient care NMDS 2022-23
- Admitted patient care NBEDS 2022-23
- Non-admitted patient NBEDS 2022-23
- Local Hospital Networks/Public hospital establishments NMDS 2022-23

*Technical Specifications*

Data set specifications for each of the relevant hospital databases can be found here: Data request specifications – Australian Institute of Health and Welfare. The technical appendices for each of the databases can be found at About the data – Hospitals – AIHW.

*Other data information – Scope of chronic conditions information in the NHDH*

In Australia's admitted patient care data, supplementary codes (ICD-10-AM-U78-U88) are used to identify chronic conditions that are present on admission but do not meet the criteria for inclusion as principal or additional diagnoses. However, these may impact on clinical care. These were introduced in July 2015 (ICD 10-AM 9th edition) to improve chronic disease data capture.

Traditionally, hospital data primarily focused on the main reason for admission and significant co-existing conditions. However, many chronic conditions, such as obesity, hypertension, or chronic kidney disease, might be present but not the primary cause for hospitalisation or require extensive treatment during the stay.

The introduction of supplementary codes in 2015 addressed this gap. They allow for a more comprehensive picture of a patient's health status, even if a chronic condition is not directly related to their hospital admission. This enhanced data capture significantly improves the following:

- Understanding chronic disease prevalence: It provides a more accurate count of individuals with various chronic conditions within the hospitalised population.

- Analysing trends: It helps in interpreting changes in diagnoses over time, as some conditions previously reported as additional diagnoses might now be captured by supplementary codes.

- Assessing multimorbidity: It is invaluable for identifying patient's multiple chronic conditions, which is essential for understanding complex health needs and resource allocation.

- Identifying specific cohorts: It allows for the identification of patient groups which chronic conditions that may not necessitate hospital care but are still part of their overall health profile.

The scope includes key disease groups such as endocrine, neurological, circulatory, respiratory, digestive, musculoskeletal, genitourinary, and congenital conditions, aligning with the supplementary ICD-10-AM code ranges (U78-U88). The above codes represent chronic conditions that are part of the patient's current health profile but are not the main reason for hospitalisation and do not require significant clinical care during the episode. Their use enhances understanding of comorbidities and overall patient health.

The supplementary codes span various disease categories, including:

U78 – endocrine, nutritional and metabolic diseases

U79 – mental and behavioural disorders

U80 – Diseases of the nervous system

U82 – Diseases of the circulatory system

U83 – Diseases of the respiratory system

U84 – Diseases of the digestive system

U86 – Disorders of the musculoskeletal system and connective tissue

U87 – Diseases of the genitourinary system

U88 – Congenital malformations, deformities and chromosomal abnormalities

These codes, part of the ICD-10-AM classification, were first included in the National Integrated Health System Information (NIHSI) 3.0 release in 2023 and subsequently in the NHDH' first release in 2024. Their inclusion did not require new data collection but rather the extraction of existing data from the National Hospitals Morbidity Data set (NHMD), which aligned with standard management practices of avoiding additional data collection burdens.

Within the NHDH's admitted patient care data set, supplementary chronic conditions are stored in the fields such as CHRON_DIAG2 to CHRON_DIAG15. Each entry is a 5-character code, combining a condition onset flag and the diagnosis code. This is consistent with ICD-10-AM coding standards.

For example, a code like "2U833" in CHRON_DIAG2 indicates:

- "2" (Condition Onset Flag): The condition did not arise during the current hospital admission. Other flags include "1" (onset during admission) and "9" (not reported).
- "U833" (Diagnosis Code): This refers to a specific supplementary chronic condition, with the decimal point omitted (e.g., U83.3, which represent asthma without mention of chronic obstructive pulmonary disease).

This structured approach ensures the NHDH provides a robust and comprehensive view of chronic conditions affecting hospitalised patients, significantly enhancing the analytical capabilities for health researchers and policymakers.

In summary, the NHDH contains information on chronic conditions via supplementary ICD-10-AM codes, covering broad range of disease groups. These are integrated into the admitted patient care data, structured clearly, and support comprehensive health analysis. The description provided for chronic condition code aligns with current coding standards and validated by the inclusion of these codes in the national hospital data sets.

*Further Information*

More information and links can be found on the AIHW's Hospitals website and National Hospitals data collection – Australian Institute of Health and Welfare

### 3.1.5 Aged Care Data Module

*Introduction*

Aged care data within the NHDH provides a comprehensive view of the services, supports, and health outcomes experienced by older Australians. It brings together information from multiple sources such as government funded aged care, health service use, and assessments. This enables a clearer understanding of the aged care sector and its interaction with the broader health system.

Importantly, aged care data in the NHDH not only reflects the utilisation of residential and home care services but also captures transitioning between aged care, hospitals and community health settings, providing a more holistic picture of ageing journey in Australia.

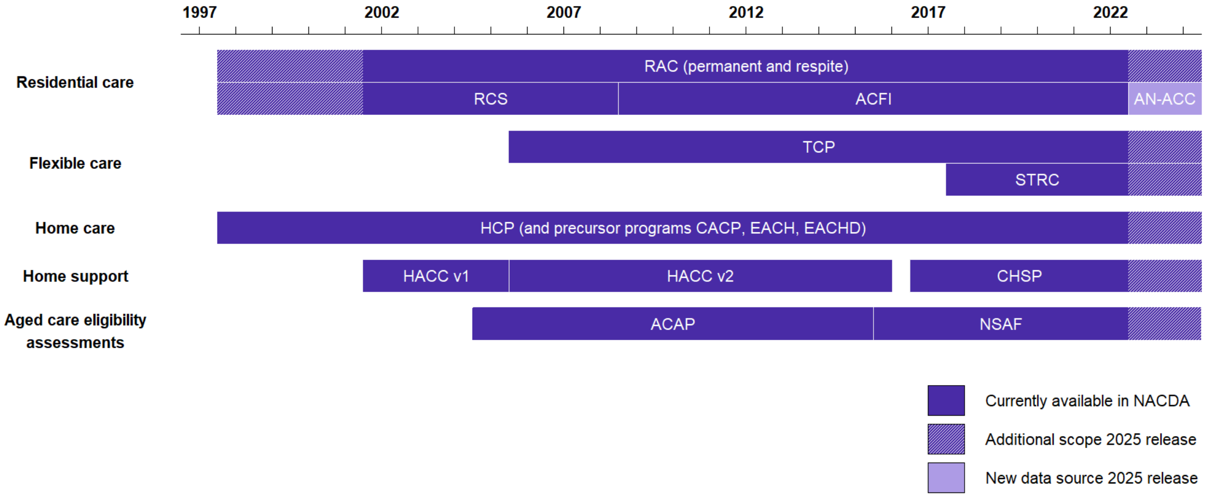
*Scope*

Aged care data in the NHDH relate to 8 different programs, each in a separate data module:

| **Care type** | **Program / data module** | **Sub types** | **Data structure** |
|---|---|---|---|
| Residential aged care (RAC) | RAC | Permanent<br>Respite<br>Funding assessments (AN-ACC, ACFI, RCS) | Periods of care with entry/exit dates |
| Home care | Home care | Home Care Packages (HCP)<br>EACH<br>EACH-D<br>CACP | Periods of care with entry/exit dates |
| Flexible | Transition Care Programme (TCP) | - | Periods of care with entry/exit dates |
| Flexible | Short-Term Restorative Care (STRC) | - | Periods of care with entry/exit dates |
| Home support | Commonwealth Home Support Programme (CHSP) | - | Single instances of services provided on particular date/s |
| Home support | Home and Community Care (HACC) | HACC v1<br>HACC v2 | Quarterly summary of service occasions |
| Eligibility assessment | National Screening and Assessment Form (NSAF) | Home Support<br>Comprehensive | Generally one-off assessment prior to care entry |
| Eligibility assessment | Aged Care Assessment Program (ACAP) | (Comprehensive only) | Generally one-off assessment prior to care entry |

* These programs migrated from System for the Payment of Aged Residential Care (SPARC); AN-ACC = Australian National Aged Care Classification; ACFI = Aged Care Funding Instrument; RCS = Residential Classification Scale; EACH = Extended Aged Care at Home; EACH-D = Extended Aged Care at Home Dementia; CACP = Community Aged Care Program.

Aged care data released in NHDH in 2024 included data to 30 June 2022. The NHDH 2025 release will include aged care data to 30 June 2024. The temporal scope of each aged care data module is summarised below, both as a figure and in text form (table).



*Source*

Aged care data within the NHDH are derived principally from the National Aged Care Data Clearing House (NACDC) – Australia's official central repository for administrative data on government-funded aged care.

The NACDC compiles anonymised, person-based records of care episodes, eligibility assessments, and provider and service details. Its input originates from multiple operational systems, including:

- Aged Care Management Payment Systems (ACMPS), capturing Home Care Packages, residential care, and flexible care programs (TCP, STRC, historical programs like CACP, EACH, EACHD)
- SPARC, used until 2022 for residential and flexible care data
- AN-ACC, for residential care classification (post-2022)
- Date Exchange (DEX), for Commonwealth Home Support Programme (CHSP) services
- My Aged Care, for eligibility assessment data (NSAF) and care-level allocations
- National Approved Provider System (NAPS)/Government Provider Management System (GPMS), detailing service and provider information

This yearly supply of data is subject to AIHW's standardisation, validation, and cleansing processes, and is then made available to researchers via the National Aged Care Data Asset (NACDA)—a continuously updated, de-identified, person-level aged care dataset. The NACDA is then integrated within the NHDH, enabling comprehensive linkage across various NHDH data assets.

More detailed information about each of the aged care programs listed above, including the source of the data and their collection and supply, is available in the National Aged Care Data Clearinghouse: User Guide.

*Data reference period*

| Care type | Program / data module | 2024 release | 2025 release |
|---|---|---|---|
| RAC | RAC | 1 July 2001 to 30 June 2022 (P, R)<br>1 July 2001 to 19 March 2008 (RCS)<br>20 March 2008 to 30 June 2022 (ACFI) | 1 July 1997 to 30 June 2024 (P, R)<br>1 Oct 1997 to 19 March 2008 (RCS)<br>20 March 2008 to 30 Sep 2022 (ACFI)<br>1 Oct 2022 to 30 June 2024 (AN-ACC) |
| Home care | Home care | 1 July 1997 to 30 June 2022 | 1 July 1997 to 30 June 2024 |
| Flexible | TCP | 1 Oct 2005 to 30 June 2022 | 1 Oct 2005 to 30 June 2024 |
| Flexible | STRC | 1 March 2017 to 30 June 2022 | 1 March 2017 to 30 June 2024 |
| Home support | CHSP | 1 July 2016 to 30 June 2022 | 1 July 2016 to 30 June 2024 |
| Home support | HACC | 1 July 2001 to 1 April 2005 (HACC v1)<br>1 July 2005 to 1 April 2015 (HACC v2) | (Not updated – historical program) |
| Eligibility assessment | NSAF | 1 July 2015 to 30 June 2022 | 1 July 2015 to 30 June 2024 |
| Eligibility assessment | ACAP | 1 July 2004 to 30 June 2015 | (Not updated – historical program) |

*How aged care content data are linked to the EMS*

There are 3 current linkages, and 3 historical linkages, of aged care identifiers to the Enhanced Medicare Spine (EMS), and there-by provision of content data for the NHDH.

- Current linkages:
  - ACMPS (RAC, HCP, TCP, STRC) use probabilistic name-based linkage methods.
  - NSAF use probabilistic name-based linkage methods.
  - CHSP use multi-step SLK581 key-based linkage methods.
- Historical linkages:
  - SPARC (historical RAC, TCP, STRC) used probabilistic name-based linkage methods.
  - HACC used multi-step SLK581 key-based linkage methods.
  - ACAP used multi-step SLK581 key-based linkage methods.

*How aged care content data modules are updated*

Keep identifiers separate to content from data custodians and holdings are merged using Linkage Maps supplied by the Linkage team. Aged care identifiers are routinely provided to the linkage team. The linkage team routinely link these identifiers to the Enhanced Medicare Spine (EMS). The cadence of linkage varies by identifier:

- ACMPS content data for NHDH are prepared and supplied annually. Each supply of ACMPS content data refreshes data from all previous years (i.e. from 1 July 1997).
- NSAF-EMS linkage is performed annually and NSAF content data for NHDH are prepared and supplied annually. Each supply of ACMPS content data refreshes data from all previous years (i.e. from 1 July 2015).
- CHSP-EMS linkage is performed annually and CHSP content data for NHDH are prepared and supplied annually. Unlike ACMPS and NSAF, each supply of CHSP data is only for the latest financial year.

*How often will aged care data in the NHDH be updated*

The aged care data in the NHDH are updated annually.

*Which tables will be updated, and how (resupply or new tables)*

Tables for the following historical programs will not be updated but will continue to be supplied in all future releases:

- ACAP
- HACC (v1 and v2)

The following table will not be updated in 2025 and future releases as the data are incorporated in assessment data:

- HCP_ASSESS

The following tables will not be updated in 2025 and future releases as they contain static data for that financial year:

- CHSP_EPI_2017 through to CHSP_EPI_2022
    - Note that CHSP_EPI for the forward financial years will continue to be supplied (e.g. in the 2025 release, CHSP_EPI_2023 and CHSP_EPI_2024 will be supplied).

The following tables will be updated as a one-off in the 2025 release only:

- ACFI (funding assessment within RAC) was replaced by AN-ACC on 1 October 2022. ACFI tables will be updated to cover the period to 1 July 2022 to 30 September 2022.
- RCS (funding assessment within RAC) was replaced by ACFI on 20 March 2008. However, the RCS table will be updated to incorporate data from 1997 through to 30 June 2001. This additional scope (1997 to 2001) is being applied to all relevant tables within RAC for the 2025 release.

All the other tables will be updated annually. In most cases, this means the entire table will be resupplied every year (and overwritten in the NHDH, with previous versions archived). For example, the coverage of HCP_EPI table is currently 1 Jul 1997 to 30 Jun 2022. In 2025, the entire table will be resupplied with coverage 1 Jul 1997 to 30 Jun 2024

*Other methodological considerations*

A synthetic ID was applied to several aged care data modules in the NHDH 2024 release (i.e. NSAF, HACC, ACAP) and will be applied to all aged care data modules in the future.

The synthetic ID is a NHDH-generated ID used to replace the source program client ID. A synthetic ID uniquely represents an individual only within a specific collection of tables, which is typically a single aged care program. The criteria for each program-specific synthetic ID are outlined in the 'Data variable list for the NHDH'.

The purpose of including synthetic IDs is to (a) replicate the original table structure and be able to differentiate records that belong to an individual within a program, including those that did not link to the EMS (do not have a PPN) and (b) be able to join relational tables for records that did not link to the EMS (do not have a PPN). Note that a PPN with multiple synthetic IDs suggests a unique individual that has multiple IDs in the source program data. For unlinked records (represented by a missing PPN), it cannot be reliably established whether multiple synthetic IDs refer to the same person or different individuals.

### 3.1.6 National Disability Insurance Scheme (NDIS) Data Module

*Introduction*

Since 1991, the AIHW has collected a Disability Services National Minimum Data Set, with updates over time to reflect changes in policy and service delivery. In 2009, the National Disability Agreement introduced a revised data set, which remained in use until the 2018–19 reporting period. From that point, disability services transitioned to the National Disability Insurance Scheme (NDIS), which became the primary source of national disability data. The NDIS provides individualised support to people with disability and is administered by the National Disability Insurance Agency (NDIA), a statutory agency of the Australian Government.

*Scope*

In 2024 the NDIA supplied the AIHW with five limited NDIS content tables for inclusion in the NHDH establishing the infrastructure and governance arrangements required for managing NDIS information. In 2025, a more comprehensive collection of NDIS data managed by the AIHW Disability Unit for NDIA was established. This expanded collection of tables replaces earlier versions and provide a new baseline for integration, longitudinal analysis and reporting within the NHDH.

This data enables more detailed, person-centred insights into service utilisation, equity and outcomes, thereby strengthening the evidence base for policy development and program evaluation.

*Source*

NDIS Content data are collected by the Commonwealth to administer the NDIS and covers the following components:

- Access requests: Contains history of applications for the NDIS, including dates, status of the application and decisions regarding the application.
- Early Childhood Early Intervention (ECEI) gateway: Summarises children in the NDIS (under age 9 as of 2023) who enter the Early Childhood Early Intervention system, as well as the types of support used.
- Mainstream Interfaces: Describes the intersection of other known mainstream services i.e. Other government systems providing services to the Australian public, e.g. health, mental health, education, justice, housing, child protection and employment. Most people interact with a range of services throughout their lives.
- Participant Demographics: Includes information on NDIS participants, including age, gender, primary and secondary disability types, impairment, ethnicity, and geography of residence. It may also include information on people who apply for the NDIS but have not been approved.
- Payments: Describes the history of services funded by participant’s NDIS budget, including rich detail on the type of support paid for.
- Plan Supports: Describes the history of participants’ plans and the categories of support that the participant is permitted to spend their NDIS budget on.
- Carer Demographics: Contains core demographic attributes (e.g., age, sex, relationship to participant, geography) for each registered NDIS carer.

• Early Childhood Approach (ECA): Captures ECA pathway records for children under 7, listing referral details, eligibility decisions, support outcomes and milestone dates.

• Goals: Lists each participant's NDIS plan goals and capturing goal category, progress status, and key creation/review dates.

• Participant Provider Relationships: Links NDIS participants to their service providers, capturing provider ID, relationship type (e.g., primary, plan manager), and start/finish dates.

• Plan Review Requests: Logs S48 reviews, also known as plan review requests, lodged when a participant's circumstances change or they believe their funded supports are inadequate. It captures reason, lodgement-to-decision timeline, outcome, and linked plan identifiers for each request.

• Participant Outcomes: Holds participant survey responses, self-reported measures of independence, social participation, health and economic wellbeing, captured at plan reviews to track how NDIS supports are affecting each participant over time.

• Family Outcomes table: Stores caregivers' survey responses, self-reported measures of family wellbeing, capacity to care, and community inclusion, collected at plan reviews to assess how NDIS supports impact participants' families over time.

• Outcomes Maps: A metadata reference table which catalogues all outcome-survey instruments used by the NDIS, including survey forms, question text, and allowable answer options. It contains no participant-level data and should be used to decode responses in Participants and Family Outcomes tables.

• Support Catalogue: Lists every NDIS support item, item number, description, unit of measure, location-specific price (inc. remote loadings), and validity dates, serving as the master price guide for plans and payments.

• Support Item: Records every claimed support instance under a participant's plan including item number, category, service date, delivered quantity, unit price and total claim amount. It enables granular analysis of utilisation and spends.

*How NDIS content data are linked to the EMS*

Participant records are linked to the EMS through name-based probabilistic matching in two passes, a strategy that accommodates occasional discrepancies in variables such as sex and date of birth. Where multiple NDIS participant identifiers resolve to the same individual, often because of name or address changes, manual review confirms the duplication. Researchers can identify these cases by noting instances in which several participant IDs map to a single NHDH PPN.

*How NDIS content data modules are updated*

The NDIA provides NDIS data extracts to the AIHW. Each data extracts delivers a complete refresh of all NDIS data tables. The update process involves three key steps:

- Linkage Preparation: The AIHW linkage team generates a linkage map to connect participant records to the EMS.
- Content Extraction: The AIHW Disability Unit prepares the content extracts from the NDIA data.
- Data Integration: The AIHW merging team combines the linkage and content data to update the NHDH modules.

The snapshots include a combination of new data as well as updates to data from prior quarters. Most tables are structured around circumstances that change and for these the most recent data are a complete history. There are two notable exceptions:

- The Participant Demographics, Access Request and Early Childhood tables are structured to have one record per person. Analysts using Participant Demographics and Access Requests should use historic quarters to observe longitudinal changes per individual. For example, a participant's disability can change and to examine disability as known at a prior date the historic cuts of data are needed.
- The Family Outcomes, Participants Outcomes, and Payments tables are updated but the NDIA only supply the most recent 36 months data with each quarterly supply. The NHDH cut of NDIS will be constructed such that each table will be complete.

*When NDIS data modules are updated*

Current NDIS annual update typically reflect data that are approximately six months behind the current date. Following delivery, there is an additional three-month period during which the data are prepared and ingested into the NHDH. As data management processes mature, it is anticipated that these timeframes will shorten, improving the currency of data available for reporting and research.

*Data reference period and states/ territories coverage*

The NDIS contains all states and territories data from 1 July 2017 to 30 June 2022 as of May 2025 NHDH release.

*Naming conventions for NDIS Content Data*

NDIS quarterly data sets adopt a clear naming convention: the base table name is suffixed with _YYZZQ, where YYZZ pairs the final two digits of the financial year's start (YY) and end (ZZ) to indicate the full fiscal span, e.g., 2425 denotes the 2024-25 financial year and Q (1-4) specifies the exact quarter. Thus, NDIS_F_OUT_24252 represents the Family Outcomes extract for the second quarter of FY 2024-25. This uniform scheme keeps each quarterly snapshot distinct while preserving lineage across reporting periods, simplifying automated ingestion, archiving, and longitudinal analysis.

*Other Methodological considerations*

Person ID is a synthetic NHDH-generated scramble of the source program's original NDIS client identifier. Because the same scrambling algorithm is applied to every table in the program, the synthetic ID travels intact through assessments, events, services and any other module tables. It is therefore the key you turn to whenever a record has not linked to the EMS and so carries no Person-Project Number (PPN). The PPN aggregates one or more synthetic IDs that the linkage process has judged to belong to the same real-world individual and promotes that person to a hub that can be connected across multiple programs or out to other NHDH data sets. In other words, the synthetic ID maintains coherence within the module, while the PPN extends that coherence across topics.

Researchers should be aware that the original five NDIS content tables delivered in 2024 differ structurally from the expanded set of tables delivered in 2025. Changes include updated table names, revised variable names, and expanded content coverage. As a result, direct comparisons or longitudinal analyses across the two versions may be misleading without careful reconciliation. The 2025 expanded delivery should be treated as a separate new delivery of data, representing a new baseline for NDIS data within the NHDH. Users are advised to consult metadata and data dictionaries for each delivery to ensure accurate interpretation and integration.

*Further Information*

More information can be found on the Data Sets | NDIS

### 3.1.7 Australian Immunisation Register (AIR) Data Module

*Introduction*

The Australian Immunisation Register (AIR) is a national register that records vaccines given to all people in Australia. The AIR includes vaccines given:

- Under the National Immunisation Program
- Through school programs and privately, such as influenza or travel vaccines.

Originally established as the Australian Childhood Immunisation Register in 1996 to record data on vaccinations provided to children aged up to 7 years, the register was expanded in 2016 to include data for people of all ages. As of 2018 the HPV vaccination register was transferred to the AIR along with records of other school-based immunisation programs. Reporting of adult vaccinations to the AIR was mandated in July 2021 for vaccinations provided under the National Immunisation Program, for influenza and for covid-19, and in December 2022 for Japanese encephalitis virus vaccines. Reporting of other vaccinations is not mandatory.

*Scope*

AIR tables included in the NHDH contain information on:

- Demographic information including sex, Indigenous status and state of residence from individuals' most recent Medicare card as recorded at the time of extract
- Vaccination episodes recorded in AIR that were administered before age 16
- Vaccination episodes recorded in AIR that were administered on or after age 16
- Details of medical contraindications to vaccines
- Natural Immunity details for specific diseases
- Vaccination supply details
- Antigen Codes and Vaccine Codes
- Information on COVID-19 vaccine code, recommended dose, minimum and maximum intervals between doses, overdue days, scheduled start and end date.

*Source*

The AIR is operated by the Services Australia under the *Australian Immunisation Register Act of 2015*. The Department of Health, Disability and Ageing (the Department) is the data custodian. The AIHW accesses AIR data from the Department's Enterprise Data Warehouse.

*How AIR content data are linked to the EMS*

The AIR data are linked to the EMS using Personal Identification Numbers (PINs). Synthetic IDs are also created for the individuals not eligible for Medicare or where their Medicare Numbers were not available.

*How are the AIR content data modules updated*

Keep identifiers separate to content from data custodians and holdings are merged using Linkage Maps supplied by the AIHW Linkage team.

*When AIR data modules are updated*

Content data modules are updated annually.

*Data reference period and states/ territories coverage*

As of May 2025, NHDH release, the AIR contains all states and territories data from 1 July 2010 to 30 June 2022 for Children data set and data from 1 July 2018 to 30 June 2022 for adults' data set

*Naming conventions for AIR Data*

The vaccination episodes data module has been split into child and adult tables based on age at the time of vaccination. Vaccine episodes administered before age 16 have been included in the child table (AIR_CONTENT_VACCNTN_EPSD_CHILD). Vaccine episodes administered on or after age 16 have been included in the adult table (AIR_CONTENT_VACCNTN_EPSD_ADULT).

All other NHDH AIR data modules are named in accordance with the Department of Health, Disability and Ageing AIR data feed specifications.

Detailed information on the variables included in these AIR data tables is available in the NHDH data item list.

*Other Methodological considerations*

The following link provides some methodological considerations when analysing AIR data for linkage projects:

Australian Immunisation Register: Technical notes – Australian Institute of Health and Welfare

### 3.1.8 National Death Index (NDI) Data Module

*Introduction*

The National Death Index (NDI) is a Commonwealth database that contains records of deaths registered in Australia since 1980.

*Source*

Data come from Registrars of Births, Deaths and Marriages in each jurisdiction, the National Coronial Information System and the Australian Bureau of Statistics (ABS). NDI records are supplemented with cause of death information using a once-off data linkage with the National Mortality Database (NMD). This enhancement enables research that requires both fact of death (whether a person died) and cause of death (what the person died from).

*Scope*

Current NDI tables included in NHDH contain NDI records from 1 July 1997 to 31 December 2023. The NDI also includes data on Indigenous status from 1998 onwards.

*How NDI content data are linked to the EMS*

The NDI content data are linked to the MCD using names, dates of birth and death and addresses as a probabilistic name-based linkage. The spine is created based on the linkage results and includes components from the NDI for the linked individuals.

*How NDI content data modules are updated*

Keep identifiers separate to content from data custodians and holdings are merged using linkage maps supplied by the AIHW linkage team.

*When NDI data modules are updated*

AIHW receives monthly updates from the jurisdictional registries, but they are not immediately reflected in the NDI module in NHDH. NDI updates are linked to the spine every quarter. NDI module in NHDH is updated annually to reflect the new causes of death updates provided by ABS.

*Data reference period and states/ territories coverage*

The NDI contains all states and territories data from 1 July 1997 to 31 December 2023 as of May 2025 NHDH release.

*Naming conventions for NDI data module*

NDI data module has been split into a content table and a by-exception table module as follows:

- NDI_CONTENT
- NDI_CONTENT_BYEX

Detailed information on the variables included in this NDI data module is available in the NHDH data item list

*Other Methodological considerations*

ABS releases are in August (preliminary Cause of Death (CoD) for previous year) and December (revised and finalised data from one and two years ago) every year. The updates are reflected in the following NHDH NDI module releases.

*Further information*

More information can be found on the Stats National Mortality Database

### 3.1.9 Australian and New Zealand Intensive Care Society (ANZICS) Data Module

*Introduction*

The Australian and New Zealand Intensive Care Society (ANZICS) is a professional organisation that supports and promotes intensive care medicine. Their data registries, managed by the ANZICS Centre of Outcome and Resource Evaluation (CORE), collect clinical data from various intensive care units (ICUs) across Australia and New Zealand. These registries provide critical information for monitoring patient outcomes, evaluating treatment effectiveness, and improving the quality of intensive care services.

*Source*

The ANZICS Adult Patient Database (APD) and ANZICS Paediatric Intensive Care Registry (PICR) are key components of this data collection. These registries capture detailed information on patient demographics, treatments received in the ICU, severity of illness scores, and outcomes (e.g., mortality, length of stay).

*Scope*

The ANZICS content data collection in the NHDH will only include Australian data, with New Zealand data not being received as part of this collection.

*How ANZICS content data are linked to the EMS*

ANZICS data are linked via SLK-581, as well as the following variables to support linkage activities:

- Postcode of Treatment Facility
- State of Treatment Facility
- Fact of death (where available)
- Date of death (where available)

*How ANZICS content data modules are updated*

ANZICS provides an annual data extract to AIHW.

*When ANZICS data modules are updated*

ANZICS data are updated annually, with updates scheduled each March.

*Data reference period and states/ territories coverage*

The ANZICS contains all states and territories data from 1 January 2017 to 30 June 2024 for Adult Patient Database (APD) and data from 1 January 1997 to 30 June 2024 Australian and New Zealand Paediatric Intensive Care Registry (ANZPICR) as of May 2025 NHDH release.

*Naming conventions for ANZICS data module*

ANZICS data tables have been split into a content table and a by-exception table module as follows:

- ANZICS_ANZPICR
- ANZICS_ ANZPICR_BYEX
- ANZICS_ANZPICR_ADX
- ANZICS_APD
- ANZICS_APD_BYEX

Detailed information on the variables included in these ANZICS data tables is available in the NHDH data item list

*Further information*

More information about ANZICS data is available from the Data Access and Publication Policy - ANZICS

### 3.2 Frequency and approximate release times for NHDH data modules

The table below provides an overview of the core data sets available through the NHDH. It outlines the frequency with which each data set is updated and the most recent release date of data availability. This information supports users in understanding the timeliness and periodicity of data releases for informed analysis and reporting.

| Data sets | Periodicity | Level | NHDH data last updated |
|---|---|---|---|
| Medicare Benefits Schedule (MBS) | Quarterly | Person | August 2025 |
| Pharmaceutical Benefits Scheme (PBS) | Quarterly | Person | August 2025 |
| Admitted patient care data drawn from the National Hospital Morbidity Database (NHMD) – public hospitals (including chronic conditions) | Annual | Person | Jul 2025 |
| Admitted patient care data drawn from the NHMD – private hospitals (including chronic conditions) | Annual | Person | Jul 2025 |
| National Non-Admitted Patient Emergency Department Care Database (NNAPEDCD) | Annual | Person | Jul 2025 |

| | | | |
|---|---|---|---|
| National Non-Admitted Patient episode-level Databases (NNAP(el)D) | Annual | Person | Jul 2025 |
| National Death Index (NDI) | Quarterly | Person | May 2025 |
| Australian Immunisation Register (AIR) | Annual | Person | May 2025 |
| National Aged Care Data Clearinghouse (NACDC) | Annual | Person | Dec 2024 |
| National Disability Insurance Scheme data set (NDIS) | Annual | Person | Oct 2024 |
| Australian and New Zealand Intensive Care Society (ANZICS) | Annual | Person | March 2025 |

## 4. Data Update Strategy

The NHDH Data Refresh Strategy is described as below:

- **Case-by-Case Assessment:**
  The data update approach for each NHDH data module is determined on a case-by-case basis in consultation with data custodians, domain experts, and other relevant stakeholders. This ensures that each data refresh aligns with the specific needs and goals of the business and Data Custodians.

- **Update Approach:**
  There are two separate approaches for data updates: Appending and Rolling. The choice of approach depends on the nature of the data and their intended use. The key considerations when deciding which option to choose are as follows:
  - **Appending Updates:**
    When new data represent an addition to an existing data set (e.g., new records or observations without altering previous data within a data set), appending updates are used. In this case, no year suffix or timestamp is typically added to the data set. For example, this approach is used for updating PBS, MBS data modules in the NHDH.

    *Exception:* This approach may not be applicable in scenarios where the table structure changes significantly between refresh cycles (e.g., new columns, changed schema). In these cases, additional planning and coordination with stakeholders is required to ensure consistency across data versions. The rolling updates approach could be used in such cases.
  - **Rolling Updates:**
    When a data set represents a time-specific snapshot (e.g., period-based reports, monthly metrics), the "rolling" strategy is used. This approach involves appending a timestamp or a unique time-based suffix (e.g., year, month) to each new data set and retaining previous data sets within the NHDH analytical environment. This allows for retention of information essential for tracking changes over time, ensuring the historical accuracy of each data set added to NHDH using this approach. It may also be preferable where table structure

changes significantly between updates. For example, this approach is used for updating patient demography modules in the NHDH.

# 5. Useful references and resources for NHDH data

## 5.1 Data Quality Statement

Data Quality Statements (DQS) for the NIHSI and NHDH provide information on linkage rates for the different modules and sub-modules. The NHDH DQS is not yet available; however, the data quality statements for the corresponding NIHSI data sets are currently accessible. More information on the data quality statement is available in the following data quality statements:

- National Integrated Health Service Information (NIHSI) version 1.0
- National Integrated Health Service Information (NIHSI) version 2.0
- National Integrated Health Service Information (NIHSI) version 3.0

The NIHSI v 1.0 data set was completed in August 2021 and made available to AIHW analysts. In December 2021. The NIHSI AA v 1.0 held data from 1 July 2010 to 30 June 2019.

The NIHSI v 2.0 data set was released in December 2022 and made available to approved analysts. The NIHSI v 2.0 held data from 1 July 2010 to 30 June 2020.

The NIHSI v 3.0 data set was released in March 2023 and made available to approved analysts. The NIHSI v 3.0 held data from 1 July 2011 to 30 June 2021.

## 5.2 Data item list

Data item lists for all data modules in the NHDH are available as a spreadsheet on the Data & data items page of the NHDH website. The spreadsheet contains worksheets for data variables across all NHDH data sets, including default and by exception data modules.

## 5.3 Technical notes on the NIHSI and NHDH

NHDH technical notes are available on the NHDH website (Resources - Australian Institute of Health and Welfare). The notes contain more detailed information about the data sources, data coverage, data linkage and data quality used in the NIHSI and NHDH.

## 5.4 Other resources

All NHDH support resources are available on the NHDH website National Health Data Hub - Australian Institute of Health and Welfare

# 6. Data Support

If users have any general questions or require assistance related to data available in the NHDH, please contact the nhdh@aihw.gov.au.

# Appendix

## Glossary

| | |
|---|---|
| National Health Data Hub (NHDH) | AIHW National Health Data Hub (NHDH) repository which contains hospital services data, Medicare Benefits Schedule, (MBS), Pharmaceutical Benefits Scheme (PBS) data, National Aged Care Services data, National Death Index (NDI) data, National Disability Insurance Scheme data set (NDIS), Australian and New Zealand Intensive Care Society (ANZICS) and Australian Immunisation Register (AIR) |
| Australian Institute of Health and Welfare (AIHW) Data Custodian | AIHW Data Custodian is an AIHW staff member with delegation from the AIHW Chief Executive Officer to exercise overall responsibilities for a specified data collection in accordance with legislation, the AIHW's governance instruments and any specific conditions for use applicable to that data collection. |
| Third party | An individual or organisation external to the participant and non-participant jurisdictions as represented on the NHDH Advisory Committee, and the AIHW. |
| Jurisdiction | A state or territory (as represented by their respective or the Australian Government (as represented by a Department). |
| AIHW Human Research Ethics Committee (HREC) | The AIHW HREC ensures that data involving individuals are managed ethically and responsibly. It reviews the ethical acceptability of health and welfare projects and collections are managed in line with national ethical standards. |
| Person Project Number (PPN) | Person Project Number is assigned for each unique individual. A PPN is unique for each person across data tables in the NHDH. A record will only be allocated a PPN if the record has linked to the EMS.<br>A record without a PPN indicates the record has not linked to the EMS. |
| Content Data | This refers to all data in a collection independent of identifiers and excluding lookup tables or other non-person related data. |
| Identifiers | This refers to information that identifies individuals, which is commonly used in the context of data linkage. |
| Person Identifier Number (PIN) | This is a series of digits that is uniquely associated with an individual within the Medicare Consumer Directory. |
| SLK | Statistical linkage key. A key that enables two or more records belonging to the same individual to be brought together. |
| Secure Access Environment (SAE) | A SAE is a restricted area that has controlled settings designed to protect sensitive data by implementing strict access and security measures. This includes Research Only Network (RON), Secure Environment for Analysing Data (SEAD), Integration Authority (IA), and Secure Unified Research Environment (SURE). |

| | |
|---|---|
| Secure Environment for Analysing Data (SEAD) | SEAD is a secure access environment that is hosted by the Australian Bureau of Statistics (ABS). |
| Research Only Network (RON) | RON is a secure access environment that is hosted by AIHW. |
| Data Item List | A data item list includes data that will be provided to the AIHW for inclusion in the NHDH. |