



Australian COVID-19 linked data set

The COVID-19 (coronavirus) pandemic has led to substantial individual, health system, and broader social and economic effects that will continue to evolve. The pandemic has also demonstrated the need for better national health data infrastructure, such as disease registers, to support evidence-based public health policy decisions.

The AIHW analyses and publishes evidence of the medium-term and long-term effects of many health conditions and diseases, including COVID-19.

In April 2022, the AIHW was funded by the Medical Research Future Fund to establish a national linked data platform using existing health data sets to strengthen evidence-based public health and health system planning and management for current and future pandemics.

Linking COVID-19-related data sets provides new insights into the health outcomes for people who have been diagnosed with the disease, and the effect that COVID-19 has had on the health system and broader community. It also offers researchers the opportunity to explore a range of issues associated with the pandemic.

Quick facts



The data linkage project combines COVID-19 case data with a range of other data sets, including hospitals and deaths data.



The linked data can be used to look at the effects of COVID-19 on health outcomes and health service usage.



Researchers can use the data set to explore issues such as re-infection rates and the effect of vaccines and treatments.

What is data linkage?

Data linkage is a process that combines information from multiple sources about the same person or entity to create a new combined data set. The results of the linkage can provide a richer story than would be possible from a single data source.

The COVID-19 linked data set is enabling us to provide a fuller picture of the relationship between COVID-19 and risk factors, the best approaches to prevention or early intervention, and the effectiveness of health and safety interventions. This will give a more holistic picture of the connections between COVID-19 and broader medical and social wellbeing.



What information is included?

Federal, state and local government agencies, as well as health and welfare service providers, hold an enormous amount of administrative and clinical data.

Information on COVID-19 cases is sourced from participating states and territories as well as the Australian Government's National Notifiable Disease Surveillance System, and a number of other administrative data sources.

The sources that are currently included are:



state/territory notifiable diseases data: COVID-19 cases



Medicare Benefits Schedule (MBS)



National Notifiable Disease Surveillance System (NNDSS)



Pharmaceutical Benefits Scheme (PBS)



Medicare Consumer Directory (MCD)



National Hospital Morbidity Database: admitted patient care data (NHMD)



Australian Immunisation Register (AIR)



National Non-Admitted Patient Emergency Department Care Database: emergency department presentations (NNAPEDCD)



National Death Index (NDI)



National Aged Care Data Clearinghouse: aged care data. (NACDC)

The register will soon include the following additional datasets:



National Disability Insurance Scheme (NDIS)



Australian and New Zealand Intensive Care Society (ANZICS)

How is your privacy protected?

The AIHW manages this data with respect for its sensitivity, and with privacy and confidentiality assured through legislation, accountability practices and procedures. The AIHW also manages relationships with data custodians to ensure accountability and appropriate use of the relevant data collections.

The AIHW protects the privacy of an individual through a process of de-identification. This involves removing identifying information (for example, a person's name, address or Medicare number) so that researchers are unable to tell who the information belongs to.

Researchers are only allowed to publish data that has been approved by the data custodian and checked that there is no way a person can be identified.

This project has ethical approval from:

- the AIHW Ethics Committee (EO2021-2-1232)
- the NSW Population and Health Services Research Ethics Committee- as part of the National Mutual Acceptance Scheme (2021_ETH00412) and
- the Northern Territory Department of Health and the Menzies School of Health Research (2021-4106).

The AIHW uses the Five Safes framework to reinforce management of the privacy and confidentiality of data. Five Safes is an approach to assessing and managing risks associated with data sharing and release. Under the framework, the risk of re-identification is minimised, particularly as data are supplied to researchers in secure access environments where outputs are checked by the AIHW.



Using the data for research

Researchers who wish to use the data will need to ensure their research question/s falls under the approved uses. Examples of approved broad areas of research are described in Table 1.

Researchers will also need to complete a project proposal form, which will be reviewed by the AIHW custodian and approved by the relevant data custodians.

For more information, contact covid19register@aihw.gov.au or visit <https://www.aihw.gov.au/reports-data/covid-linked-data-set>.

Table 1. Approved research themes and potential analysis questions

RESEARCH THEME	POTENTIAL ANALYSIS
Epidemiological and statistical research	<ul style="list-style-type: none"> • Estimates of COVID-19 incidence and prevalence • Burden of disease analysis • Cause and timing of death after a COVID-19 diagnosis • Associations between comorbidities, vaccination and risk factors for people who have died from COVID-19 • Estimates of COVID-19 severity • Estimates of re-infection with COVID-19 • Impact of COVID-19 vaccines and treatments
Service use and medication dispensing and patient journeys	<ul style="list-style-type: none"> • Patterns of health and aged care service use and expenditure before and after COVID-19 diagnosis • Patterns of medication dispensing before and after COVID-19 diagnosis • Impact of COVID-19 on long-term service needs
Identifying groups or cohorts of interest	<ul style="list-style-type: none"> • Impact and extent of infection in key settings of interest, for example in the health and aged care settings • People with chronic conditions and whether COVID-19 diagnosis is associated with changes in chronic condition outcomes, treatment and management • Health and service use outcomes for particular patient cohorts such as those with long COVID, older Australians, those with comorbidities or who speak a language other than English as their main language at home, those who have received a COVID-19 vaccine
Monitoring, evaluation and data quality improvement	<ul style="list-style-type: none"> • Monitoring and evaluation of policies and programs designed to reduce the transmission of COVID-19 • Monitoring and evaluation of COVID-19 vaccination and treatments (such as antivirals) • Verification of information from the National Notifiable Disease Surveillance System (NNDSS)