

Appendix B. Data quality and limitations

The relevance and validity of research undertaken to analyse trends and to model the factors associated with targets is highly dependent on the quality and timeliness of available data when this report was prepared. The main issues identified across the datasets are:

- the quality of identification of First Nations people
- the possible undercounting of First Nations people
- the change in Indigenous identification over time.

As well, the modelling analyses are limited by the availability and quality of measures in the data, as well as by small numbers of the First Nations population. Many important factors measuring individuals' socioeconomic background, connection to culture and Country, experience of racism and discrimination, access to culturally safe service and so on, are often not available in the survey data used for modelling analysis in this report.

It must also be noted that the associations observed in the modelling outcomes using survey data should not be interpreted as causation. All data used in statistical modelling in this report are cross-sectional data – collected at one point in time – and do not allow the timeline or the order of life events to be identified or change over time to be assessed.

In addition there are known limitations specific to each dataset that need to be taken into consideration when interpreting the results of the analyses. Table B.1 presents a list of the data sets used in this report and a summary of known issues related to each.

Table B.1: List of data sources and limitations

Data set	Data limitations
2021 Australian Census of Population and Housing	Infrequent collection and non-response. Further information on data quality can be found at https://www.abs.gov.au/census/guide-census-data/census-methodology/2021/quality-declaration .
Australian Census Linked Dataset 2011-2016	False links can occur during the linkage process: even when a record pair matches on all or most linking fields, it may not actually belong to the same individual. Sampling error is possible due to a small proportion of the total population being used to produce estimates that represent the whole population. Further information on data quality can be found at https://www.abs.gov.au/ausstats/abs@.nsf/Previousproducts/2080.0Quality%20Declaration02011-2016?opendocument&tabname=Notes&prodno=2080.0&issue=2011-2016&num=&view= .
National Mortality Database 2017–2021	Not all First Nations deaths are correctly identified in death registrations. The level of First Nations mortality is likely to be underestimated. Further information on data quality can be found at https://www.abs.gov.au/methodologies/deaths-australia-methodology/2022 and https://www.abs.gov.au/methodologies/causes-death-australia-methodology/2022 .

(continued)

Table B.1 (continued): List of data sources and limitations

Data set	Data limitations
AIHW National Perinatal Data Collection, 2018–2021	Quality of Indigenous identification – underreporting of First Nations data; small numbers in certain maternal conditions and baby outcomes can create variability in time series. Further information on data quality can be found at https://meteor.aihw.gov.au/content/799298 .
Australian Early Development Census, 2021	Quality of Indigenous identification; subjective measurement of child and child’s family related items. Further information on data quality can be found at https://www.aedc.gov.au/resources/about-the-aedc .
National Early Childhood Education and Care Collection (NECECC) 2022	The current methodology results in enrolment rates for early childhood exceeding 100% for some jurisdictions. Further information on data quality can be found at https://www.abs.gov.au/ausstats/abs@.nsf/Lookup/4240.0.55.001Chapter62013 .
ABS Schools data 2022	Occasionally, different jurisdictions make changes in the administration of their education system that can impact on the coherence of the statistics over time. The comparability of these statistics may vary between state and territories, and that of schooling sectors, where different policies and administrative arrangements may affect different outcomes. Further information on data quality can be found at https://www.abs.gov.au/methodologies/schools-methodology/2021 .
2018–19 National Aboriginal and Torres Strait Islander Health Survey	Survey results are subject to sampling errors – a small proportion of the population is used to produce estimates that represent the whole population. Non-sampling errors may include: non-response, misunderstanding of questions or errors in recording, coding or processing the survey; responses being possibly affected by imperfect recall or individual interpretation of survey questions. Further information on data quality can be found at https://www.abs.gov.au/statistics/people/aboriginal-and-torres-strait-islander-peoples/national-aboriginal-and-torres-strait-islander-health-survey/latest-release .
2021 National Prisoner Census	Indigenous status is self-identified with Indigenous status not known for fewer than 1% of prisoners in 2022. Further information on data quality can be found at https://www.abs.gov.au/methodologies/prisoners-australia-methodology/2022

(continued)

Table B.1 (continued): List of data sources and limitations

Data set	Data limitations
Youth Justice National Minimum Data Set, 2022–23	The data are subject to revisions each year and may not match previously published data. The most recent data are the most accurate. About 5% of all young people in the Youth Justice National Minimum Data Set since 2000–01 have an unknown Indigenous status. Further information on data quality can be found at https://meteor.aihw.gov.au/content/772562 .
AIHW collection of the Child Protection National Minimum Data Set, 2022–23	Data produced from the Child Protection National Minimum Data Set are based on nationally agreed specifications and may not match state and territory figures published elsewhere; they also may not be comparable with data for previous years. There are differences in jurisdictional policy, practice, legislation and data systems that can affect the interpretation of child protection data. Further information on data quality can be found at https://meteor.aihw.gov.au/content/757571 .
National Hospital Morbidity Database, 2018-2021	Interpretation of results should take into account the relative quality of the data from the jurisdictions. Time series analyses may be affected by changes in the quality of Indigenous identification over time. Time series presentations in this report include data from both public and private hospitals across several jurisdictions, so the overall effect of changes in Indigenous identification over time is unclear. Further information on data quality can be found at https://www.aihw.gov.au/reports-data/myhospitals/content/about-the-data .
2014–15 National Aboriginal and Torres Strait Islander Social Survey	Survey results are subject to sampling errors – a small proportion of the population is used to produce estimates that represent the whole population. Non-sampling errors may include non-response; misunderstanding of questions or errors in recording, coding or processing the survey; responses being possibly affected by imperfect recall or individual interpretation of survey questions. Further information on data quality can be found at https://www.abs.gov.au/ausstats/abs@.nsf/PrimaryMainFeatures/4720.0?OpenDocument .