

TITLE PAGE

Title:

Data Resource Profile: Seeding Success: a cross-sectoral data resource for early childhood health and development research in Australian Aboriginal and non-Aboriginal children

Author list:

Kathleen Falster^{*1,2,3}, Mikaela Jorgensen³, Mark Hanly³, Emily Banks^{1,2}, Marni Brownell⁴, Sandra Eades⁵, Rhonda Craven⁶, Sharon Goldfeld^{7,8}, Deborah Randall³, Louisa Jorm³.

***Corresponding author:**

Dr Kathleen Falster

Postal address: The Sax Institute, PO Box K617, Haymarket, NSW, Australia, 1240

Email: kathleen.falster@anu.edu.au

Phone: +612 9188 9536, Fax: +612 9188 9501

Author affiliations:

¹National Centre for Epidemiology and Population Health, Research School of Population Health, Australian National University, Canberra, Australia.

²The Sax Institute, Sydney, Australia.

³Centre for Big Data Research in Health, UNSW Australia, Sydney, Australia.

⁴Manitoba Centre for Health Policy, University of Manitoba, Winnipeg, Canada.

⁵Baker IDI Heart and Diabetes Institute, Melbourne, Australia.

⁶Institute of Positive Psychology and Education, Australian Catholic University, Sydney, Australia.

⁷Murdoch Children's Research Institute, Melbourne, Australia.

⁸Department of Paediatrics, University of Melbourne, Melbourne, Australia.

Keywords

Indigenous population, child development, inequality, child health, cohort studies, Australia, data linkage, vulnerable populations.

Word count

3145 words

Data Resource Basics

Background

In common with other indigenous populations worldwide,(1) Aboriginal and Torres Strait Islander Australians experience worse health, development and later life outcomes than non-Indigenous Australians across most routinely reported population measures throughout the life course.(2-5) These inequalities are founded in a history of colonisation and discriminatory legislation, policy and practices.(6) Despite the Australian Government's commitment in 2008 to 'Closing the Gap' in outcomes between Indigenous and non-Indigenous people, progress towards the majority of targets has been largely disappointing to date.(7) It is well recognised that the social conditions of children and their families are an important determinant of their health and development and the early childhood years represent an opportunity for interventions that may improve health and development trajectories.(8, 9) Accordingly, a number of longitudinal cohort studies have been established in recent years to better understand the child, family and community factors that promote positive and negative health and development outcomes in Aboriginal and Torres Strait Islander children.(10-13) Large-scale, population-level epidemiological research is also required to identify policy sensitive targets for intervention, and to assess their impact. To conduct this research, longitudinal data resources with large-scale population coverage are needed.

Purpose and scope of the Seeding Success data resource

The Seeding Success data resource was established to conduct child health and health equity research in Australia's most populous state, New South Wales (NSW), with a focus on Aboriginal and Torres Strait Islander children and scope to investigate multiple dimensions of disadvantage. It is a unique resource comprising individual-level administrative data with comprehensive population coverage, routinely collected by several sectors (including health, education, and community services), that have been linked together to construct the health and development trajectories of a large, population-based cohort of children from birth to school age. More details about the rationale, aims, and methodological plan are available elsewhere.(14) In brief, the Seeding Success data resource capitalises on recently available data from the Australian Early Development Census (AEDC; formerly the Australian Early Development Index(15)), a population level measure of child development at school entry, linked to other administrative data sources to enable investigation of: (i) early life characteristics that promote positive and negative early childhood development; (ii) how early childhood outcomes vary geographically and how features of local communities contribute to this variation; and (iii) the relation of child outcomes to program and service delivery.

The Seeding Success data resource setting: New South Wales, Australia

Australia is divided into five state and two territory jurisdictions (Figure 1). The state of NSW is home to almost one-third of Australia's population (7.2 million of 22.3 million people) and nearly one-third of Australia's Aboriginal and Torres Strait Islander population (209 000 of 670 000 people).(16) Aboriginal and Torres Strait Islander people comprise 2.9% of the total NSW population; of these, 95% identify as Aboriginal, 3% as Torres Strait Islander, and 2% as Aboriginal and Torres Strait Islander.(16) Because Aboriginal people are the original inhabitants of NSW(17) and account for 95% of the NSW Indigenous population, the term 'Aboriginal' will be used throughout this paper.(18) The majority of the NSW population lives in major cities (74%), followed by regional (26%) and remote (<1%) areas. In contrast, 44% of Aboriginal people in NSW live in major cities, 51% in regional areas, and 5% in remote areas.(19)

The Seeding Success data resource population

The Seeding Success data resource includes a population-based cohort of children who were born in NSW and started school in 2009 (N=79 432) or 2012 (N=86 846) (Table 1). Births were identified from the NSW Register of Births, Deaths and Marriages and the NSW Perinatal Data Collection; birth dates ranged from January 1 2002 – December 31 2008. School starters were identified from the 2009 and 2012 AEDC, which has high population coverage in NSW (99.1% in 2009 and 97.3% in 2012)(2, 20).

In total, data are available for 166 278 children. Data on children's parents are also available. For 163 590 children, data are available for their mother, and for 38 878 children, data are also available for their 'second' parent, of whom 99.7% are fathers (Table 1). Of the total 166 278 children in the cohort, 154 935 children (93.2%) were born and started school in NSW, and 11 343 children (6.8%) were born in NSW and started school in another jurisdiction (Figure 2).

Data Collected

Data sources and measures

Data were provided by national and state government agencies. Table 1 summarises the individual-level measures available in the Seeding Success data resource for children and their parents, from the administrative data sources described below. Figure 3 illustrates the time coverage of the measures available from each data source. For children in the data resource, data were obtained from all sources. Where data was obtained for the parents, this is indicated below.

The *NSW Perinatal Data Collection (PDC)* includes records for all children born at ≥ 20 weeks gestation or weighing ≥ 400 grams in NSW public or private hospitals, as well as planned home births. It includes

demographic variables, and information on maternal health, the pregnancy, labour, birth, and perinatal outcomes. For the mothers of the children in the resource, the month and year of birth from each of her prior birth records in the PDC were obtained to enable calculation of the mother's age at first motherhood.

The *NSW Registry of Births, Deaths and Marriages* compiles birth registrations for NSW. Birth registrations include month and year of birth and Aboriginal status for mothers and other parents (mostly fathers).

The *NSW Register of Congenital Conditions* includes records of congenital conditions identified during pregnancy, at birth or during the first year of life, as well as the date of diagnosis. Because identifiers are removed from the register after five years, data were only available for linkage for pregnancy outcomes recorded in the period 2004-2008.

The *NSW Admitted Patients Data Collection* includes records of all public and private hospital separations (discharges, transfers, and deaths) in NSW since 1 July 2001. It includes patient demographics, diagnoses and procedures coded according to the Australian Modification of the International Statistical Classification of Diseases and Related Problems, 10th revision (ICD-10-AM).(21) For parents of children in the resource, hospitalisation records for the five years prior to the child's birth were obtained, where available.

The *NSW Emergency Department Data Collection* includes records of all presentations to metropolitan, and the majority of regional, emergency departments in NSW since 1 January 2005. It includes patient demographics, mode of arrival, triage category, mode of separation, diagnoses, and procedures coded according to ICD-10-AM(21) or Systematized Nomenclature of Medicine, Clinical Terms (SNOMED CT®).

The *NSW Mental Health Ambulatory Data Collection (MH-AMB)* includes the assessment, treatment, rehabilitation or care of non-admitted patients since January 2000, although there was significant undercounting of contacts until 2005/6. The MH-AMB data collection includes patient demographics, diagnosis codes coded according to the ICD-10-AM,(21) and other characteristics of the service provided for each 'contact' between a clinician and a patient. For parents of children in the resource, MH-AMB records for the five years prior to the child's birth were obtained, where available.

The *Key Information Directory System (KiDS)* includes records of all child protection contacts with the NSW Department of Family and Community Services (FACS) since 2003, including information about whether a child has: 1) been assessed by a caseworker as being at actual harm/risk of harm; 2) had a legal decision made in relation to them; 3) been placed in out-of-home care; or 4) been referred to and participated in a FACS early intervention program.

The *Australian Early Development Census* is a population measure of child development that has been collected nation-wide every three years since 2009 for children enrolled in their first year of formal full-time school.(15) In Australia, the school year starts in late January/early February and the majority of children

start school at age five. The AEDC is a teacher-completed checklist, collected between May and August, and includes items about the child's development on five domains: 1) physical health and well-being; 2) social competence; 3) emotional maturity; 4) language and cognitive skills; and 5) communication skills and general knowledge. At the time data were linked for Seeding Success, AEDC data for 2009 and 2012 were available.

NSW Public School Enrolment data includes demographic information about the child and his or her family, including parent education and occupation, for children enrolled in NSW Public Schools (i.e. government funded schools). In NSW in 2009 and 2012, 70% of children in their first year of school were enrolled in a Public School, 20% in a Catholic school and 10% in an Independent school.(22)

For children in the data resource, approval has also been obtained to link to: (1) Medicare Benefits Schedule data, which consists of records for claims for medical and diagnostic services; and (2) Centrelink income assistance data, consisting of records of receipt of Australian Government payments for families with low incomes. Details of these linkages have been described elsewhere;(14) they have not yet commenced.

Where standard units of geography for areas of residence are available in the resource (e.g. PDC birth or AEDC record), publically available area-level information, such as geographical remoteness or socioeconomic indices, have been attached to child and parent records.

Data linkage

The NSW Centre for Health Record Linkage (<http://www.cherel.org.au/>) linked the individual-level data from the sources described above to create the Seeding Success data resource. The Australian Institute of Health and Welfare (AIHW) will undertake the planned linkages of Medicare and Centrelink data.

The majority of data sources currently in the resource, with the exception of the KiDS and the school enrolment data, are routinely linked within the Centre for Health Record Linkage Master Linkage Key, which includes a set of regularly updated links within, and between, core population data sources in NSW.

Custodians of each data source provide the Centre for Health Record Linkage with an encrypted source record number and demographic details (including full name, address, date of birth, sex) for each record in the source data; in the case of data sources not in the Master Linkage Key, this is done on an ad-hoc basis. Records for each individual are then linked within and between data sources using probabilistic methods based on demographic details.(23)

For each individual identified in the linkage process for this resource, a project person-specific number (PPN) was created; this PPN was then assigned to all records that belonged to that individual in each data source. Following linkage, the PPN and associated source record numbers for each data source were returned to the

relevant data custodians. The data custodians then extracted and supplied the approved variables and relevant PPNs to the Study Investigators. In this way, the separation principle was applied, such that no-one working with the data had access to both identifying information and the content data.(24)

Measurement of Aboriginality using multiple, linked population data sources

Because Aboriginal people are known to be under-enumerated in administrative data,(25-29) and use of multiple sources of linked data has been shown to enhance Aboriginal enumeration in population studies,(30-33) we derived the child's Aboriginality from multiple data sources available for children at birth and school age. The Aboriginality of the child may be derived from the Aboriginal status recorded for the mother, second parent (mostly fathers) or child, depending on the source data, or a combination from multiple data sources. Table 2 summarises the number of children recorded as Aboriginal in the birth and school age source data, and the enhanced enumeration of Aboriginal children when Aboriginality is derived from multiple linked data sources. Of the child cohort, 10 430 children (6.3%) were classified as Aboriginal using the 'ever identified' identification algorithm (i.e. the child or their parents were recorded as Aboriginal on any of the available birth records, or the AEDC) (Table 2), which was applied for the descriptive statistics presented in Table 3.

Table 3 presents select characteristics for Aboriginal and non-Aboriginal children in the cohort, and their parents, available from one or more of the linked data sources. Aboriginal children were more likely to be born to mothers who were unmarried, did not have private health insurance, lived in more socioeconomically disadvantaged areas, and were younger than mothers of non-Aboriginal children. Although the majority of mothers of Aboriginal children lived in major cities or inner regional areas at the time of their child's birth, mothers of Aboriginal children were more likely to live in outer regional and remote/very remote areas compared with mothers of non-Aboriginal children (20.3% versus 7.8%, and 6.0% versus 0.7%, respectively). Based on teacher report on the AEDC, 4.1% of Aboriginal children were repeating their first year of school compared to 2.3% of non-Aboriginal children in the cohort, and medically diagnosed special needs were more common among Aboriginal children compared with non-Aboriginal children (7.4% versus 4.5%).

Ethical approval and study governance

The study has received ethical approval from the NSW Population and Health Services Research Ethics Committee (AU RED Reference: HREC/14/CIPHS/23, Cancer Institute NSW reference: 2014/04/523), the NSW Aboriginal Health and Medical Research Council Ethics Committee (1031/14), the Australian Institute of Health and Welfare Ethics Committee (EO2015/2/141) and the Australian National University Human

Research Ethics Committee (2014/384). A Reference Group consisting of Aboriginal community organisations, service providers and their representatives, has been established to provide the Study Investigators with guidance on the study aims, priorities and interpretation of the findings, to advise on community engagement, and to connect with relevant groups and organisations to facilitate translation of findings into policy and practice.

Funding

Funding from the National Health and Medical Research Council (NHMRC) (Project grant number 1061713) enabled creation of the Seeding Success data resource.

Data Resource Use

The Seeding Success data resource is currently being analysed to investigate the relationships between child development outcomes and maternal age at childbirth, gestational age at birth, preschool attendance, and school starting age (relative to NSW school enrolment policy), in Aboriginal and non-Aboriginal children. The resource is also being used to examine injury-related emergency department visits and hospitalisations during early childhood, and child development at age five, for children in contact with the child protection system in NSW, including children whose families participated in an early intervention program (Brighter Futures) that aims to prevent families from entering, or escalating within, the child protection system. Children whose families were exposed to the early intervention program have been identified in the data resource, as well as a propensity score matched comparison group of unexposed children who were similar on available measured characteristics. All analyses will be written up for publication in peer-reviewed scientific journals and/or reports to relevant policy agencies.

Strengths and Weaknesses

A significant strength of this resource is the use of population data with high coverage and large numbers of children. As such, statistical power has been optimised and selection bias minimised. Moreover, these data will enable the experience of Aboriginal children, a small and vulnerable population group, to be made visible. The use of statewide, linked population data will also enable exploration of geographic variation in outcomes across metropolitan, regional and remote locations. The linkage of data about program participation to this resource is another strength, and will enable the application of quasi-experimental methodologies to assess the impact of program participation on early childhood outcomes.

Some key strengths and limitations of the main outcome measure in the resource – child development, measured by the AEDC – should also be acknowledged. A major strength is the availability of the AEDC for

an almost complete population of school starters in census years. Although teachers complete the AEDC checklist about individual children, it is not designed as a diagnostic tool. Rather, the AEDC is a holistic population measure of child development that can be used to identify groups of children, or communities, that may benefit from prevention and early intervention to improve child outcomes, including school readiness.(15) A limitation of the AEDC as an outcome measure includes bias that might be introduced via teacher completion of the checklist, which could be affected by the characteristics of the teacher, the child, the school, or a combination of factors. Another challenge is the suitability of the instrument for use with culturally and linguistically diverse groups of children. However, the AEDC and its predecessor, the Canadian Early Development Instrument,(34) have been the subject of numerous studies that suggest acceptable measures of validity and reliability, and similar psychometric properties of the instrument in different countries.(34-39) In Australia, there have also been studies examining the appropriate use and interpretation of the AEDC with Indigenous and culturally and linguistically diverse communities.(40-42) An important limitation of this resource is that Aboriginal people are less likely to have a birth registration(27), are known to be under-enumerated in administrative datasets,(25, 26, 28, 29) and the recording of Aboriginal status has changed over time in some data sources.(25, 29) However, use of multiple linked data sources has been shown to enhance enumeration of Aboriginal people,(30-33) which is an option available in this resource. Another limitation is that emergency department and congenital conditions data were not available for the entire study period for all children with an AEDC record in 2009 or 2012. Additionally, emergency department presentations in regional and remote areas were not complete; in 2012, it was estimated that 88% of all presentations to public hospital emergency departments in NSW were captured.(43) Service use (e.g. hospitalisations) may be underestimated for children who live in areas close to the state border because these children may access some services in the neighbouring jurisdiction. It is also not possible to ascertain from the data currently in the resource whether, or when, cohort children lived outside of NSW for any period between birth and school age. Additional linkage to national Medicare data may provide information about child mobility between jurisdictions during the study period.

Data Resource Access

The Seeding Success Investigators welcome researchers to make contact regarding collaborative project proposals and applying for data resource access. There are a few conditions of data access that must be met. First, the proposed project needs to relate to the overall program of research. The Investigators will review project proposals and determine whether specific scientific questions can be addressed using the data resource, and lie within the scope of data custodian and ethical approvals.(14) Second, the data

resource is currently stored within the Secure Unified Research Environment (SURE), which is a remote-access computing environment accessible over encrypted Internet and Australian Academic and Research Network connections.(44) Because of the highly confidential nature of the data, all data preparation and analysis must be conducted within the SURE virtual project workspace, where security controls are in place to protect the privacy and confidentiality of the data. Before researchers can obtain access to the data, researchers who will directly access the data must complete the online SURE training module, activate a SURE user account including payment of user fee, and the names of collaborating researchers must be added to the study protocol for ethical approval.

Seeding Success in a nutshell

- The Seeding Success data resource was established to conduct child health and health equity research in New South Wales (NSW), Australia, with a focus on Aboriginal and Torres Strait Islander children and scope to investigate multiple dimensions of disadvantage.
- Data is available for all children who were born in NSW, and were enrolled in their first year of full-time school and had an Australian Early Development Census record in 2009 or 2012 (N=166,278). Health data on children's parents prior to the child's birth are also available (data on mothers for 163,590 children, data on fathers for 38,878 children).
- Probabilistic data linkage was used to combine individual and longitudinal administrative data with comprehensive population coverage from several sectors, including health, education and community services, enabling child health and development trajectories to be followed from birth to school age.
- The use of multiple data sources linked at an individual-level enables enhanced identification of Aboriginal and Torres Strait Islander children who are usually under-enumerated in administrative data sources, and frequently lost to follow-up in cohort or survey studies.
- Researchers are welcome to contact the authors regarding collaborative project proposals and access to the data, which are stored in a secure, remote-access computing environment.

References

1. Anderson I, Robson B, Connolly M, Al-Yaman F, Bjertness E, King A, et al. Indigenous and tribal peoples' health (The Lancet–Lowitja Institute Global Collaboration): a population study. *Lancet*. 2016;9(388):131-57. doi: 10.1016/S0140-6736(16)00345-7. Epub 2016 Apr 20.
2. Australian Government. A Snapshot of Early Childhood Development in Australia 2012 - Australian Early Development Index (AEDI) National Report. Australian Government. Canberra: Australian Government; 2012.
3. Australian Health Ministers' Advisory Council. Aboriginal and Torres Strait Islander health performance framework 2014 report. Canberra: AHMAC; 2014.
4. Australian Institute of Health and Welfare. The health and welfare of Australia's Aboriginal and Torres Strait Islander people: an overview 2011. Cat. no. IHW 42. Australian Institute of Health and Welfare. Canberra: AIHW; 2011.
5. Australian Institute of Health and Welfare. National Partnership Agreement on Indigenous Early Childhood Development: second annual report on health performance indicators. Cat. no. IHW 151. Canberra: AIHW; 2015.
6. Brown N. History, law, and policy as a foundation for health care delivery for Australian indigenous children. *Pediatr Clin North Am*. 2009;56(6):1561-76.
7. Australian Government. Closing the gap: Prime Minister's report 2014. Canberra: Australian Government; 2014.
8. Pillas D, Marmot M, Naicker K, Goldblatt P, Morrison J, Pikhart H. Social inequalities in early childhood health and development: a European-wide systematic review. *Pediatr Res*. 2014;76(5):418-24. doi: 10.1038/pr.2014.122.
9. Marmot M, Bell R. Fair society, healthy lives. *Public Health*. 2012;126 Suppl 1:S4-10.
10. Buckskin M, Ah Kit J, Glover K, Mitchell A, Miller R, Weetra D, et al. Aboriginal Families Study: a population-based study keeping community and policy goals in mind right from the start. *Int J Equity Health*. 2013;12(1):41.
11. Comino E, Craig P, Harris E, McDermott D, Harris M, Henry R, et al. The Gudaga Study: establishing an Aboriginal birth cohort in an urban community. *Aust N Z J Public Health*. 2010;34(s1):S9-S17.
12. SEARCH Investigators. The Study of Environment on Aboriginal Resilience and Child Health (SEARCH): study protocol. *BMC Public Health*. 2010;10(1):287.
13. Thurber KA, Banks E, Banwell C. Cohort profile: footprints in time, the Australian longitudinal study of Indigenous children. *Int J Epidemiol*. 2015;44(3):789-800.
14. Falster K, Jorm L, Eades S, Lynch J, Banks E, Brownell M, et al. What factors contribute to positive early childhood health and development in Australian Aboriginal children? Protocol for a population-based cohort study using linked administrative data (The Seeding Success Study). *BMJ Open*. 2015;5(5):e007898.
15. Brinkman SA, Gregory TA, Goldfeld S, Lynch JW, Hardy M. Data resource profile: the Australian early development index (AEDI). *Int J Epidemiol*. 2014;43(4):1089-96.

16. Australian Bureau of Statistics. Estimates of Aboriginal and Torres Strait Islander Australians, June 2011. Canberra: ABS, 2013. (ABS Cat. No. 3238.0.55.001.).
17. NSW Health. Communicating Positively: A guide to appropriate Aboriginal terminology. Sydney: NSW Health; 2004.
18. Australian Bureau of Statistics. 2075.0 - Census of Population and Housing - Counts of Aboriginal and Torres Strait Islander Australians, 2011 Canberra: ABS; 2012.
19. Australian Bureau of Statistics. TableBuilder 2011 Census 2015 [Available from: <http://www.abs.gov.au/websitedbs/censushome.nsf/home/tablebuilder?opendocument&navpos=240>].
20. Centre for Community Child Health and Telethon Institute for Child Health Research. A Snapshot of Early Childhood Development in Australia: the AEDI National Report 2009. Australian Government. Canberra: Centre for Community Child Health and Telethon Institute for Child Health Research; 2009.
21. National Centre for Classification in Health. International Statistical Classification of Diseases and Related Health Problems, 10th Revision, Australian Modification (ICD-10-AM), Australian Classification of Health Interventions (ACHI) and Australian Coding Standards (ACS). Sydney, Australia: National Centre for Classification in Health; 2006.
22. Australian Bureau of Statistics. 4221.0 - Schools, Australia, 2015. Canberra: ABS; 2016.
23. Centre for Health Record Linkage. Quality assurance. <http://www.cherel.org.au/quality-assurance> (accessed 23 August 2016).
24. Kelman CW, Bass AJ, Holman CD. Research use of linked health data--a best practice protocol. Aust N Z J Public Health. 2002;26(3):251-5.
25. Australian Institute of Health and Welfare. Indigenous identification in hospital separations data - quality report. Health Services Series no. 35. Cat. no. HSE 85. Canberra: Australian Institute of Health and Welfare; 2010. Contract No.: Cat. no. HSE 85.
26. Bentley JP, Taylor LK, Brandt PG. Reporting of Aboriginal and Torres Strait Islander peoples on the NSW Admitted Patient Data Collection: the 2010 data quality survey. NSW Public Health Bull. 2012;23(1-2):17-20.
27. Gibberd AJ, Simpson JM, Eades SJ. No official identity: a data linkage study of birth registration of Aboriginal children in Western Australia. Aust N Z J Public Health. 2016;40(4):388-94.
28. Lawrence D, Christensen D, Mitrou F, Draper G, Davis G, McKeown S, et al. Adjusting for under-identification of Aboriginal and/or Torres Strait Islander births in time series produced from birth records: using record linkage of survey data and administrative data sources. BMC Med Res Methodol. 2012;12:90.
29. Thompson SC, Woods JA, Katzenellenbogen JM. The quality of Indigenous identification in administrative health data in Australia: insights from studies using data linkage. BMC Med Inform Decis Mak. 2012;12(1):1.
30. Christensen D, Davis G, Draper G, Mitrou F, McKeown S, Lawrence D, et al. Evidence for the use of an algorithm in resolving inconsistent and missing Indigenous status in administrative data collections. Australian Journal of Social Issues. 2014;49(4):423-43.

31. Gialamas A, Pilkington R, Berry J, Scalzi D, Gibson O, Brown A, et al. Identification of Aboriginal children using linked administrative data: Consequences for measuring inequalities. *J Paediatr Child Health*. 2016;52(5):534-40.
32. Randall DA, Lujic S, Leyland AH, Jorm LR. Statistical methods to enhance reporting of Aboriginal Australians in routine hospital records using data linkage affect estimates of health disparities. *Aust N Z J Public Health*. 2013;37(5):442-9.
33. Taylor LK, Bentley J, Hunt J, Madden R, McKeown S, Brandt P, et al. Enhanced reporting of deaths among Aboriginal and Torres Strait Islander peoples using linked administrative health datasets. *BMC Med Res Methodol*. 2012;12:91.
34. Janus M, Offord D. Development and psychometric properties of the Early Development Instrument (EDI): A measure of children's school readiness. *Can J Behav Sci*. 2007;39(1):1-22.
35. Brinkman SA, Silburn S, Lawrence D, Goldfeld S, Sayers M, Oberklaid F. Investigating the validity of the Australian early development index. *Early Educ Dev*. 2007;18(3):427-51.
36. Guhn M, Gadermann A, Zumbo BD. Does the EDI measure school readiness in the same way across different groups of children? *Early Education and Development*. 2007;18(3):453-72.
37. Janus M, Brinkman SA, Duku EK. Validity and Psychometric Properties of the Early Development Instrument in Canada, Australia, United States, and Jamaica. *Social Indicators Research*. 2011;103(2):283-97.
38. Muhajarine N, Puchala C, Janus M. Does the EDI Equivalently Measure Facets of School Readiness for Aboriginal and Non-Aboriginal children? *Social Indicators Research*. 2011;103(2):299-314.
39. Janus M, Brinkman S, Duku E, Hertzman C, Santos R, Sayers M, et al. *The Early Development Instrument : population-based measure for communities : a handbook on development, properties and use*. Canada: Offord Centre for Child Studies; 2007.
40. Goldfeld S, O'Connor M, Mithen J, Sayers M, Brinkman S. Early development of emerging and English-proficient bilingual children at school entry in an Australian population cohort. *International Journal of Behavioral Development*. 2014;38(1):42-51.
41. Silburn S, Brinkman S, Ferguson-Hill S, Styles I, Walker R, Shepherd CC. *The Australian Early Development Index (AEDI) Indigenous Adaptation Study*. Perth: Curtin University of Technology and Telethon Institute for Child Health Research; 2009.
42. Goldfeld S, Mithen J, Barber L, O'Connor M, Sayers M, Brinkman S. *The AEDI Language Diversity Study Report*. Melbourne: Centre for Community Child Health, The Royal Children's Hospital, Murdoch Childrens Research Institute. 2011.
43. Australian Institute of Health and Welfare. *Emergency department care 2014-15: Australian hospital statistics*. Health services series no. 65. Cat. no. HSE 168. Canberra: AIHW; 2015.
44. The Sax Institute. *SURE (Secure Unified Research Environment)*. Sydney: The Sax Institute. <https://www.saxinstitute.org.au/our-work/sure/> (accessed 15 April 2016).

Collaborators

The Seeding Success Investigator team comprises Louisa Jorm, Kathleen Falster, Sandra Eades, John Lynch, Emily Banks, Marni Brownell, Rhonda Craven, Kristjana Einarsdóttir, Deborah Randall, Sharon Goldfeld, Alastair Leyland, Elizabeth Best and Marilyn Chilvers.

Contributors

KF and LJ had overall responsibility for the conception of this study with scientific input from the chief investigators. MJ prepared the first draft of this paper, with input from KF, MH and LJ. KF wrote subsequent drafts of the paper, with input from all authors. MH and MJ prepared and analysed the data. All authors approved the final draft.

Conflict of Interest

None declared.

Funding

This work was supported by an Australian National Health and Medical Research Council (NHMRC) Project Grant (#1061713). KF was supported by an NHMRC Early Career Fellowship (#1016475) and an NHMRC capacity building grant (#573122). EB was supported by an NHMRC Senior Research Fellowship (#1042717). SE was supported by an NHMRC Career Development Fellowship (#1013418). MB was supported by the Manitoba Center for Health Policy Population-Based Child Health Research Award. SG was supported by an NHMRC Career Development Fellowship (#1082922).

Acknowledgements

The authors would like to thank the Australian Government Department of Education, the NSW Ministry of Health, the NSW Register of Births, Deaths and Marriages, the NSW Department of Education, and the NSW Department of Family and Community Services, for allowing access to the data. They also thank the Australian Institute of Health and Welfare Data Integration Services Centre, Medicare Australia, and the Australian Government Departments of Health, Human Services and Social Services for negotiating data access and linkage processes for the stage 2 linkage. The authors thank the NSW Centre for Health Record Linkage for conducting the linkage of data sources

detailed in this paper. The authors acknowledge comments on the draft manuscript from Dr Emily Klineberg (NSW Health), in addition to named collaborators.

Table 1. Number and percentage of children in the Seeding Success data resource (n=166,278) who have records from each data source, by AEDC collection year.

Data source	AEDC collection year				Source data information	
	2009 ¹		2012		Years of data linked	Key measurements
	n	%	n	%		
Children's linked data						
Australian Early Development Census	79,432	100.0	86,846	100.0	2009, 2010 ¹ , 2012	Developmental outcomes in first year of school; demographic details; preschool attendance; health and development needs.
Birth record in NSW (total)	79,432	100.0	86,846	100.0	-	-
Perinatal Data Collection (PDC) ²	78,742	99.1	85,821	98.8	01/2002-12/2008	Items relating to pregnancy, labour, baby's condition, mother's previous pregnancies.
Registry of Births, Deaths and Marriages birth registrations	78,713	99.1	85,758	98.7	01/2002-12/2008	Date of birth; demographic items for mother, father and baby at child's birth.
Register of Congenital Conditions ³	423	0.5	1,352	1.6	01/2004-01/2008	Details of congenital conditions detected during pregnancy, birth, first year of life.
Admitted Patients Data Collection ⁴	78,665	99.0	86,141	99.2	01/2002-12/2012	Hospital admissions and separations; diagnoses; procedures.
Emergency Department Data Collection ⁴	51,163	64.4	65,840	75.8	01/2005-12/2012	Mode of arrival and separation; triage category; diagnoses; procedures.
Mental Health Ambulatory Data Collection ⁴	982	1.2	1,100	1.3	04/2003-12/2012	Contacts with mental health day programs, psychiatric outpatients, outreach services; diagnoses.
Key Information Directory System ⁴	10,939	13.8	13,527	15.6	12/2002-12/2012	Items relating to contacts with Community Services, including child protection, out-of-home-care, Brighter Futures program.
Public School Enrolments ⁵	50,732	63.9	56934	65.6	2009, 2012	Mother's and father's language, occupation, schooling and tertiary education
Total cohort children with linked data for parents (prior to cohort child's birth) and siblings						
Total children with mothers who have linked data ⁶	78,265	98.5	85,325	98.2	-	-
Total children with second parents ⁷ who have linked data ⁶	14,107	17.8	24,771	28.5	-	-
Total children who have siblings ⁸ within the cohort	13,411	16.9	13,378	15.4	-	-
Total children with mothers with a linked PDC record prior to cohort child's birth					01/1994-12/2008	Mother's age; mother's date of birth; and the child's date of birth were obtained for children born to the mother of cohort child, prior to the birth of cohort child, to calculate mother's age at first birth. ⁸
Mothers	40,690	51.2	45,147	52.0		
Admitted Patients Data Collection ¹⁰					01/2001-12/2008	Hospital admissions and separations; diagnoses; procedures.
Mothers	78,133	98.4	85,269	98.2		-
Second parents ⁷	13,626	17.2	24,072	27.7		-
Mental Health Ambulatory Data Collection ¹⁰					01/2001-12/2008	Contacts with mental health day programs, psychiatric outpatients, outreach services; diagnoses.
Mothers	2,107	2.7	4,047	4.7		-
Second parents ⁷	963	1.2	2,112	2.4		-

AEDC, Australian Early Development Census; 1. Includes 1,352 children from the 2010 AEDC top-up collection; 2. PDC records contain information on both mothers and babies; 3. Because there is only ever five years of RoCC data available for linkage, few children in the 2009 AEDC collection year linked to the RoCC; 4. Children may have one or more linked records from this data source; 5. Only available for children enrolled in a NSW Public School; 6. Parent must have linked data in the PDC, APDC, or MH-AMB and child must have RBDM birth record; 7. The second parent reported on the RBDM birth registration, includes 63,228 (99.73%) males, 156 (0.25%) females, and 16 (0.3%) with missing data for sex; 8. Other children born to the same mother, identified via linkage to the RBDM, who had a 2009 or 2012 AEDC record; 9. For children born between 2002-2008, the full PDC record is available; 10. Parents may have one or more linked records from this data source, for the period five years prior to the child's birth (where available).

Table 2. Proportion and number of children identified as Aboriginal using one or more birth and school age data sources in the Seeding Success data resource.

Identification algorithm	Data source and reference person for Aboriginal status						% (n/N [‡])
	Birth Records					School Record	
	PDC (Mother)	RBDM (Mother)	RBDM (Second Parent [†])	APDC (Child)	APDC (Mother)	AEDC (Child)	
PDC Only	■						2.7 (4494/164163)
RBDM Only		■	■				4.7 (7736/164471)
APDC Only				■			2.9 (4743/166278)
AEDC Only						■	5.0 (8286/166278)
PDC or RBDM	■	■	■				5.2 (8593/166273)
Any birth record	■	■	■	■	■		5.5 (9089/166278)
Identified at least twice	■	■	■	■	■	■	4.5 (7544/166278)
Ever Identified	■	■	■	■	■	■	6.3 (10430/166278)

PDC, Perinatal Data Collection; RBDM, Register of Births, Deaths and Marriages; APDC, Admitted Patient Data Collection; AEDC, Australian Early Development Census; [†]The second parent, reported on the RBDM birth registration, includes 63,228 (99.73%) males, 156 (0.25%) females, and 16 (0.3%) with missing data for sex; [‡]Denominator varies based on combinations of datasets used and amount of missing data in source data.

Table 3. Select demographic and school entry characteristics of children in the Seeding Success data resource (n=166,278).

	Aboriginal ¹ , n (%)	Non-Aboriginal, n (%)	Total
Total n	10,430 (100.0)	155,848 (100.0)	166,278 (100.0)
Child's sex			
Male	5,313 (50.9)	80,385 (51.6)	85,698 (51.5)
Female	5,117 (49.1)	75,463 (48.4)	80,580 (48.5)
Age of mother at child's birth ²			
Median years (interquartile range)	25 (21-30)	30 (26-34)	30 (26-34)
Missing	1 (0.0)	4 (0.0)	5 (0.0)
Age of second parent ³ at child's birth ²			
Median years (interquartile range)	28 (23-33)	33 (29-37)	32 (29-37)
Missing	1,570 (15.1)	5,242 (3.4)	6,812 (4.1)
Mother's country of birth			
Australia	9,814 (94.1)	112,049 (71.9)	121,863 (73.3)
Other country	564 (5.4)	42,931 (27.5)	43,495 (26.2)
Missing	52 (0.5)	868 (0.6)	920 (0.6)
Mother's marital status at child's birth ⁴			
Married/de facto	4,909 (47.1)	129,989 (83.4)	134,898 (81.1)
Single/widowed/divorced/separated	4,742 (45.5)	21,830 (14.0)	26,572 (16.0)
Missing/unknown	779 (7.5)	4,029 (2.6)	4,808 (2.9)
Private health insurance at child's birth ⁴			
Yes	880 (8.4)	59,146 (38.0)	60,026 (36.1)
No	9,430 (90.4)	95,016 (61.0)	104,446 (62.8)
Missing	120 (1.2)	1,686 (1.1)	1,806 (1.1)
Mother's highest level of schooling			
Year 12 or equivalent	1,964 (18.8)	59,559 (38.2)	61,523 (37.0)
Year 11 or equivalent	791 (7.6)	6,549 (4.2)	7,340 (4.4)
Year 10 or equivalent	2,447 (23.5)	20,744 (13.3)	23,191 (13.9)
Year 9 or equivalent, or below	1,185 (11.4)	4,934 (3.2)	6,119 (3.7)
Missing/Not applicable	4,043 (38.8)	64,062 (41.1)	68,105 (41.0)
Second parent's ³ highest level of schooling			
Year 12 or equivalent	1,305 (12.5)	48,653 (31.2)	49,958 (30.0)
Year 11 or equivalent	505 (4.8)	5,559 (3.6)	6,064 (3.6)
Year 10 or equivalent	2,069 (19.8)	24,378 (15.6)	26,447 (15.9)
Year 9 or equivalent, or below	1,099 (10.5)	5,650 (3.6)	6,749 (4.1)
Missing/Not applicable	5,452 (52.3)	71,608 (45.9)	77,060 (46.3)
Highest occupation level of either parent ⁵			
Manager/professional	596 (5.7)	24,527 (15.7)	25,123 (15.1)
Business manager/associate professional	758 (7.3)	22,781 (14.6)	23,539 (14.2)
Trades/clerks/services	1,454 (13.9)	23,524 (15.1)	24,978 (15.0)
Driver/hospitality/labourer	1,820 (17.4)	15,256 (9.8)	17,076 (10.3)
Not in paid work in last 12 months	2,020 (19.4)	7,297 (4.7)	9,317 (5.6)
Missing/Not applicable	3,782 (36.3)	62,463 (40.1)	66,245 (39.8)
Geographic remoteness ⁶			
Major city	3,948 (37.9)	102,112 (65.5)	106,060 (63.8)
Inner regional	3,498 (33.5)	37,705 (24.2)	41,203 (24.8)
Outer regional	2,116 (20.3)	12,128 (7.8)	14,244 (8.6)
Remote/very remote	631 (6.0)	1,052 (0.7)	1,683 (1.0)
Missing	237 (2.3)	2,851 (1.8)	3,088 (1.9)
Area socioeconomic advantage and disadvantage ⁶			
First quintile (most disadvantaged)	2,179 (20.9)	12,720 (8.2)	14,899 (9.0)
Second quintile	1,921 (18.4)	16,110 (10.3)	18,031 (10.8)
Third quintile	4,211 (40.4)	52,641 (33.8)	56,852 (34.2)
Fourth quintile	1,261 (12.1)	32,413 (20.8)	33,674 (20.3)
Fifth quintile (most advantaged)	621 (6.0)	39,113 (25.1)	39,734 (23.9)
Missing	237 (2.3)	2,851 (1.8)	3,088 (1.9)
Age at AEDC collection			
< 5 years 1 month	1,828 (17.5)	27,459 (17.6)	29,287 (17.6)
5 years 1 month to 5 years 3 months	1,837 (17.6)	26,702 (17.1)	28,539 (17.2)
5 years 4 months to 5 years 6 months	2,234 (21.4)	34,999 (22.5)	37,233 (22.4)
5 years 7 months to 5 years 9 months	2,456 (23.5)	37,911 (24.3)	40,367 (24.3)
5 years 10 months to 6 years 0 months	913 (8.8)	13,130 (8.4)	14,043 (8.4)
> 6 years 0 months	1,162 (11.1)	15,647 (10.0)	16,809 (10.1)
Child is repeating school year ⁷			
No	9,815 (94.1)	149,789 (96.1)	159,604 (96.0)
Yes	427 (4.1)	3,657 (2.3)	4,084 (2.5)
Missing	188 (1.8)	2,402 (1.5)	2,590 (1.6)
Child has health and development needs			
No needs identified	6,638 (63.6)	125,342 (80.4)	131,980 (79.4)
Special needs status ⁸	772 (7.4)	6,983 (4.5)	7,755 (4.7)
Additional needs ⁹	2,388 (22.9)	16,685 (10.7)	19,073 (11.5)
Child requires further assessment	632 (6.1)	6,838 (4.4)	7,470 (4.5)

1. Defined as child or parent identified as Aboriginal on any of PDC, RBDM or APDC birth records, or AEDC school record; 2. Age derived from date of birth on RBDM birth record or PDC record where RBDM date of birth missing; 3. The second parent reported on

the RBDM birth registration, includes 63,228 (99.73%) males, 156 (0.25%) females, and 16 (0.3%) with missing data for sex; 4. Based on APDC birth record; 5. Based on highest ranking occupation of either parent; 6. Based on mother's statistical local area of residence from PDC birth record; 7. Child is repeating first year of school at the time of the AEDC; 8. Children medically diagnosed as having high needs requiring special assistance due to chronic medical, physical, or intellectually disabling conditions; 9. Children not classified as special needs status who have medically diagnosed or parent-reported physical, visual, hearing, speech, learning, emotional or behavioural problems.

Figure 1. Map of Australia with the state of New South Wales highlighted in grey.

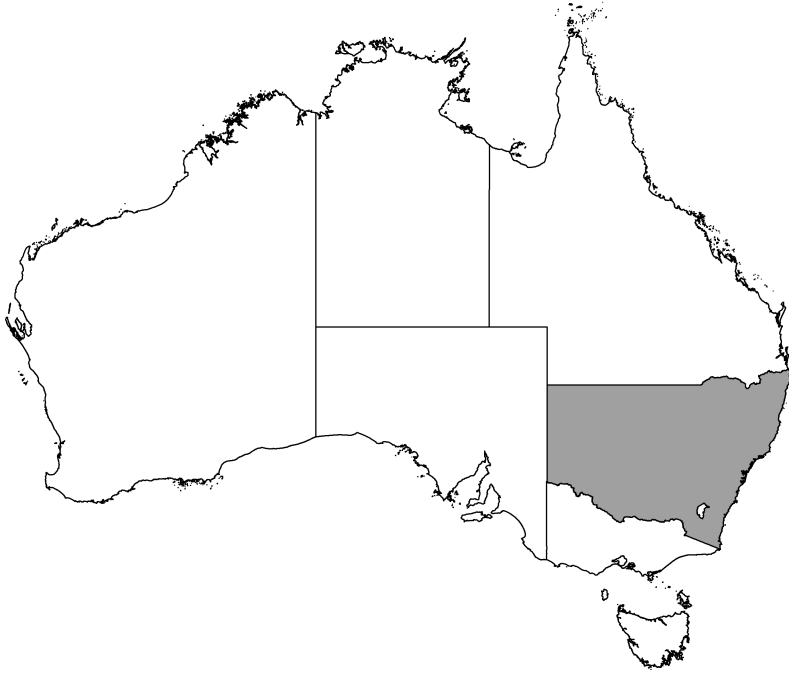
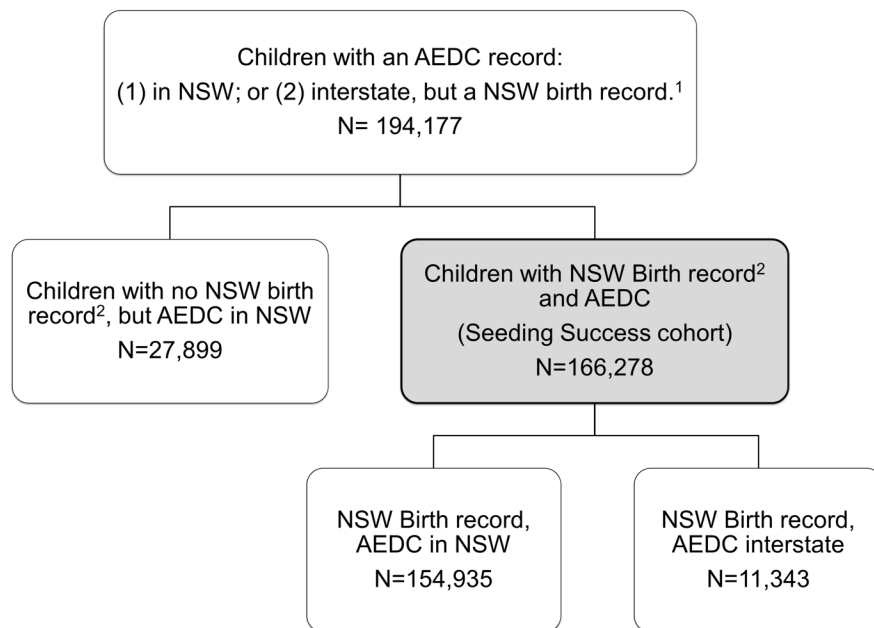


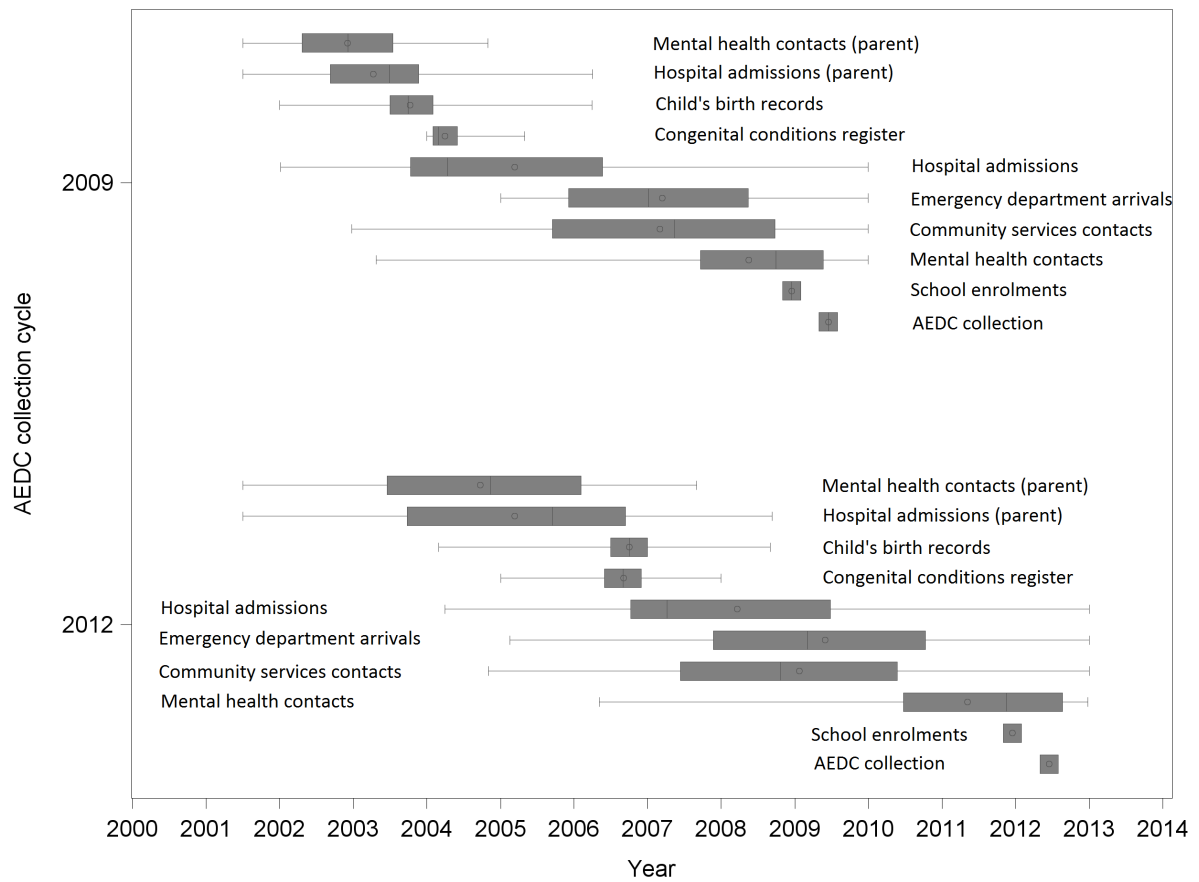
Figure 2. Process for assembling the Seeding Success data resource population from birth and school age administrative data sources



Footnote to Figure 2:

AEDC, Australian Early Development Census; NSW, New South Wales; 1. Number of children after excluding 269 non-unique birth and/or AEDC records from source data; 2. Birth records from Perinatal Data Collection and/or Register of Births, Deaths and Marriages, 2002-2008.

Figure 3. Time coverage of data available from each source, for children and their parents within the Seeding Success data resource, by AEDC collection year†



Footnote to Figure 3:

†Excludes children who participated in the 2009 AEDC top-up collection in 2010 (n=1,352; 0.8%). For each data source, the horizontal box plot represents the following descriptive statistics: circle, mean; line inside the box, median; upper edge of box, 75th percentile value; lower edge of box, 25th percentile value; endpoint of upper whisker, maximum value; endpoint of lower whisker, minimum value.