



The First Five Years: What makes a difference?

1. Introduction

The First Five Years Project is a collaborative endeavour involving project partners from across government and the university sector. The project aimed to create an enduring data asset linking a measure of childhood development (including physical health and wellbeing, social competence, emotional maturity, language and cognitive skills (school-based), and communication skills and general knowledge) with family, social, economic and health data and data about early childhood education and care attendance and quality. This enables researchers to better understand the effects family, social, economic and health events and circumstances have on early child development using integrated data held by government.

The Multi-Agency Data Integration Project (MADIP), which has since been renamed the Person Level Integrated Data Asset (PLIDA), was a partnership among Australian Government agencies to develop a secure and enduring approach for combining government datasets to create a comprehensive picture of Australia over time. MADIP was used to integrate multiple data sources for the First Five Years project, including the Australian Early Development Census (AEDC), child care administrative data, the Census of Population and Housing, tax, health and welfare data.

The project was supported by the Data Integration Partnership for Australia (DIPA), a three-year partnership across 20 Commonwealth agencies which commenced on 1 July 2017 and ended 30 June 2020.

Objectives

The First Five Years project explores rates of child developmental vulnerability and other measures of development, as reported in the Australian Early Development Census (AEDC) during the first year of a child's full-time schooling, to answer the following research questions:

- (1) What child and family characteristics contribute to children being developmentally vulnerable at age 4-6 years?

- (2) How does child care attendance affect child developmental vulnerability, and do these effects differ based on child care usage patterns and quality of services?

Data and methods

This project used a MADIP microdata extract with the safe and secure linkage of AEDC data. This project was approved by the DIPA and MADIP Boards and the data was accessed through the secure Australian Bureau of Statistics (ABS) DataLab environment. The ABS adopted the Five Safes framework which takes a multi-dimensional approach (safe people, projects, settings, data and output) to balance disclosure risk and data utility. For more information, see [Five Safes framework | Australian Bureau of Statistics \(abs.gov.au\)](#).

Using MADIP as a data source provides detailed microdata for the AEDC cohort and their parents/carers, which might otherwise only be available through extensive and invasive surveying. The creation of the First Five Years enduring data asset is the first time in the national context that AEDC outcomes have been linked to a wide range of data on student's family, income, health, employment, socio-economic and income support status.

Our research centred on a cohort of approximately 274,000 children, using data about them and their carers across their early childhood years until 2018, their first year of full-time school. The developmental vulnerability of these children in their first year of full-time school was measured using the 2018 AEDC. AEDC data are collected across five key areas of early childhood development, known as domains:

- physical health and wellbeing,
- social competence,
- emotional maturity,
- language and cognitive skills (school-based),
- communication skills and general knowledge.

The domains are sometimes grouped into cognitive domains (language and cognitive skills (school-based) and communication skills and general knowledge) and non-cognitive domains (social competence, emotional maturity, and physical health and wellbeing).

A range of characteristics of the children and their families and their social and economic contexts were obtained from AEDC questionnaire results, child care data, Census of Population and Housing, and other tax, health and welfare data. The datasets used in this project and their sources are listed in Table 1 in the *Methodology* section.

To investigate which child and family characteristics contribute to children being developmentally vulnerable at age 4-6 years, we conducted **descriptive analysis** to explore how:

- (1) rates of being developmentally vulnerable on each domain,
- (2) rates of being developmentally vulnerable on one or more domains (DV1),



(3) rates of being developmentally on track on all domains¹.

were associated with a range of child, family, and social characteristics.

Using a variety of **predictive modelling** techniques, including logistic regression and random forests, we then looked at which associations remained and were strongest in predicting the child's developmental outcomes when controlling for important factors in our model.

To understand how child care affects developmental vulnerability, we used **statistical inference modelling** to explore associations between developmental vulnerability and early childhood education attendance, quality and duration. We used statistical techniques such as G-computation and Inverse Propensity Weighting to adjust for a range of confounding variables that may bias the association between child care and preschool attendance and developmental vulnerability.

The term “early childhood education and care” (ECEC) generally refers to both preschool and child care, so the report uses “child care” when it is important to distinguish between these. Child care attendance data are drawn from Child Care Management System (CCMS) administrative data, and cover all sessions (including centre-based day care and family day care) that children used from birth up to their first year of full-time schooling in 2018. Child care quality is identified using the Australian Children's Education & Care Quality Authority (ACECQA) National Quality Standard (NQS) rating.

For more information on data and methodologies, please see the *Methodology* section.

Results

A range of child, family and social characteristics were associated with developmental vulnerability

- The analysis found that some characteristics were important predictors of developmental vulnerability in children, even after taking into account other differences in a child's circumstances.
- Higher parent or carer educational attainment, higher neighbourhood socio-economic status and higher household income were associated with lower rates of developmental vulnerability.
- Rates of developmental vulnerability were higher among children from language backgrounds other than English and those with a parent not born in an OECD country.
- Rates of developmental vulnerability were higher among children of mothers under 20 and children of single parents.
- Rates of developmental vulnerability were higher among children with parents experiencing mental ill-health.

¹ The proportion of children who were developmentally on track on all domains was calculated by dividing the number of children who were classified as developmentally on track on all five of the AEDC domains by the total number of children who had valid domain scores in all five AEDC domains. This definition is different to the AEDC OT5 indicator. For details, see the *Methodology* section.



Different child care usage patterns and quality were associated with different rates of developmental vulnerability

- Aboriginal and Torres Strait Islander children, children with language backgrounds other than English and children from low SES areas with 15 to 30 weekly hours of child care attendance had higher rates of being developmentally on track in all domains, when compared to lower or higher hours of attendance.
- Children who attended formal child care exhibited lower rates of overall developmental vulnerability (that is, they had lower rates of being developmentally vulnerable on at least one developmental domain), though some individual domains showed a different pattern.
- Children had lower rates of developmental vulnerability if they attended child care that was above standard quality (services with an overall rating of “Excellent” or “Exceeding NQS”) compared to children at lower quality services.
- Children reported to have attended preschool had much lower rates of overall developmental vulnerability than those who did not.

Statistical techniques were used to estimate the effects of child care attendance, hours, and quality on developmental vulnerability, taking into account other underlying differences between children

- Statistical techniques for inference were used to better estimate the effect of child care on developmental vulnerability by comparing children with similar socio-economic backgrounds. However, while adjusting the results to account for different circumstances gives a better estimate of effects, non-observable parameters (such as quality of home care) mean that more work is needed to infer a causal relationship.
- The effect of child care attendance depended on total duration and intensity and varied by AEDC domain:
 - Regular use of low-medium weekly hours of child care consistently resulted in lower rates of developmental vulnerability.
 - Most children attended child care at medium levels of average and total hours. This group had a lower rate of developmental vulnerability for the communication skills and general knowledge and language and cognitive skills (school based) domains.
 - Higher average and total hours of child care were associated with elevated risks of developmental vulnerabilities for the emotional maturity and social competence domains.
- Higher quality child care improved overall outcomes for children compared to lower quality child care.

Outline of reports

We report the results from our project in a series of factsheets, organised as follows:

- *Characteristics associated with developmental vulnerability in children:*



- **2.1 Child and family characteristics:** In this section we show how socio-economic factors such as parental educational attainment, household income and neighbourhood SES were associated with child developmental vulnerability. Other family composition factors such as single/dual parent household, language background and maternal age of the mother at birth are also investigated. How child and parent ill-health affected child development is presented separately for boys and girls.
- **2.2 Predictive modelling of developmental vulnerability in children:** We identify the risk and preventative factors that had a significant association with early child development, and how much each individual factor was associated with child developmental vulnerability, for boys and girls separately.
- *Association of child care and developmental vulnerability*
 - **3.1 Associations between early education attendance and child development:** In this section we describe the association between child care and preschool attendance and rates of being developmentally on track for different subgroups of the population (Aboriginal and Torres Strait Islander, language background other than English, single parent, low SES). The relationship between parental and child-centric factors and the rates of formal child care attendance is also presented.
 - **3.2 Quality of child care:** In this section we describe the associations between child care quality and rates of being developmentally on track for different subgroups of the population.
 - **3.3 Hours of child care attendance:** In this section we describe the associations between child care hours and rates of being developmentally on track for different subgroups of the population.
 - **3.4 Estimating the effect of child care on children's developmental outcomes:** In this section we estimate the effect of child care attendance and duration, and quality on developmental vulnerability using statistical inference techniques such as G-computation and Inverse Propensity Weighting to control for the differences in children's circumstances.
- *Summary and discussions*
 - **4. Conclusion**
- *Methodology and appendix*
 - **5.1 Methodology:** this section contains detailed information about datasets, key analytical decisions, constructed factors and statistical techniques used within the project.
 - **5.2 Appendix:** this section discusses linkage bias, sample size for the inference computation results, and the rules used for calculating income. It also contains analysis using 2021 AEDC data and the child care administrative data, for a comparison of charged and attended child care hours, though a full analysis using the 2021 AEDC data is not available due to the lack of linkage to other data sources.

For further information on the First Five Years project, please contact the research team at IntegratedDataAnalytics@education.gov.au

