Australian Government
Department of Education

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 domains[1].

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

Australian Government
Department of Education

The First Five Years: What makes a difference?

2.1 Child and family characteristics

Key Findings

We analysed the associations between child and family characteristics and developmental vulnerability. Guided by the literature and the characteristics available in the data, we considered parental education, single-parent status, socio-economic circumstances, language background, maternal age, and health.

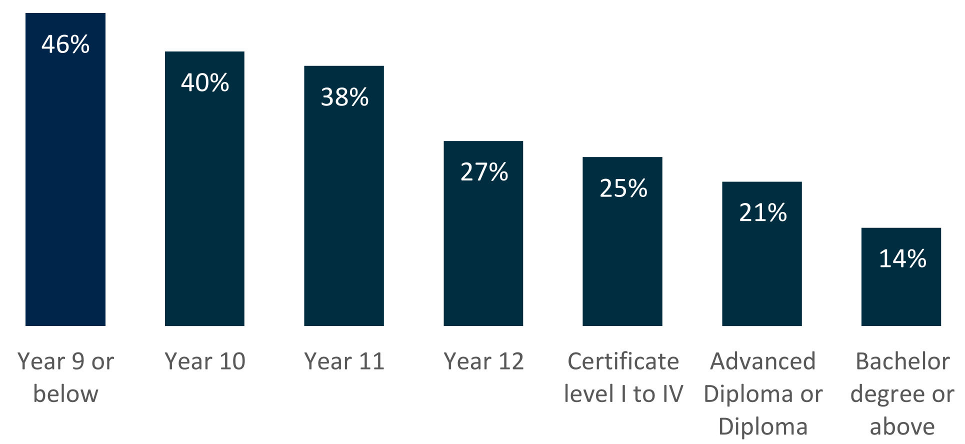
- Higher parent or carer highest educational attainment was associated with lower rates of developmental vulnerability.
- Neighbourhood socio-economic status and household income were protective factors for child development. Children of families in the third lowest household income decile had the highest overall rates of developmental vulnerability.
- Children from language backgrounds other than English and those with a parent not born in an Organisation for Economic Cooperation and Development (OECD) country were substantially more likely to be developmentally vulnerable on the communication skills and general knowledge domain.
- Rates of developmental vulnerability were higher among children of mothers under 20 and children of single parents. Children born when their mothers were aged between 28 and 39 years were least likely to be developmentally vulnerable.
- Parental mental ill-health also increased the risk of developmental vulnerability in children, particularly long-term parental mental ill-health.

Parent or carer educational attainment

The Australian Early Development Census (AEDC) reports the highest level of school and post-school education attained for up to two parents or carers[1]. We used this information to derive the highest educational attainment of the most educated and least educated parent or carer for each child (see *Methodology* section).

Overall, we found higher levels of parent or carer educational attainment were associated with lower rates of child developmental vulnerability, both for the most educated parent (Figure 1), and for each parent individually (Figure 2). These results are consistent with previous studies, which have identified family education levels as a key attribute related to developmental vulnerability, typically with a focus on maternal education (Kohl and Hobbs 1998; Lapointe et al. 2007; Noble et al. 2015; Comuk-Balci et al. 2016). Past research suggests the mechanism behind this relationship may stem from associated attributes like intellectual capability and greater income (Venetsanou and Kambas 2010; Noble et al. 2015). Additionally, previous studies indicate that parental education may contribute to child development through more stimulating environments, positive psychology and greater literacy of health and child rearing practises (Venetsanou and Kambas 2010; Comuk-Balci et al. 2016).

Figure 1. Proportion (%) of children who are developmentally vulnerable on one or more domains, by highest educational attainment of most educated parent, 2018 AEDC cohort.



Source: Customised ‘First Five Years’ extract from the Multi-Agency Data Integration Project.

Notes: This figure compares children from the 2018 cohort of the Australian Early Development Census. N = 259,034. Grouped by highest educational attainment of most educated parent or carer.

For children with two parents or carers listed in the AEDC, we found that higher educational attainment in both parents/carers positively impacted child development. Figure 2 shows how different combinations of parent or carer educational attainment were associated with the risk of child developmental vulnerability.

[1] The education details recorded in the AEDC are collected from school enrolment records, which collect data for up to two parents or carers.

Figure 2. Proportion (%) of children who are developmentally vulnerable on one or more domains, by highest educational attainment of the most educated and the least educated parents or carers, 2018 AEDC cohort.

Less-educated parent \ More-educated parent	Bachelor degree or above	Advanced Diploma or Diploma	Certificate level I to IV	Year 12	Year 11	Year 10	Year 9 or below
Bachelor degree or above	13%						
Advanced Diploma or Diploma	15%	19%					
Certificate level I to IV	15%	19%	22%				
Year 12	16%	20%	23%	22%			
Year 11	18%	24%	28%	28%	32%		
Year 10	20%	26%	29%	31%	38%	36%	
Year 9 or below	25%	32%	37%	37%	43%	47%	42%

Source: Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.

Notes: This figure compares children from the 2018 cohort of the Australian Early Development Census. N =235,149. Data only available for up to two parents or carers. Parent or carer education was identified using the AEDC (see Methodology section).

For children with two parents or carers listed in the AEDC, as the highest educational attainment of either parent or carer decreased, the proportion of children that were developmentally vulnerable on one or more domains generally increased. This trend was consistent at all education levels except as it passed through the Certificate level I to IV group when holding the less-educated parent's education level constant. This is likely because the AEDC contains a broad education level grouping Certificate I to IV, treating four distinct education levels under the Australian Qualifications Framework (AQF) as one group (DET 2019). There is a large spread of difficulty and length within the Certificate I-IV grouping reported in the AEDC, making this difficult to order with the other qualifications listed.

The importance of the less educated parent's or carers' education can be illustrated by looking at the last row in Figure 2. For a parent or carer with a highest educational attainment of Year 9 or below, even if the other parent or carer has a Bachelor degree or above, the rate of child developmental vulnerability on one or more domains was 25 per cent. As the level of highest educational attainment of the more educated parent or carer decreased, the percentage of children that were developmentally vulnerable on one or more domains increased, with rates at 47 per cent and 42 per cent for Year 10 and Year 9 or below, respectively.

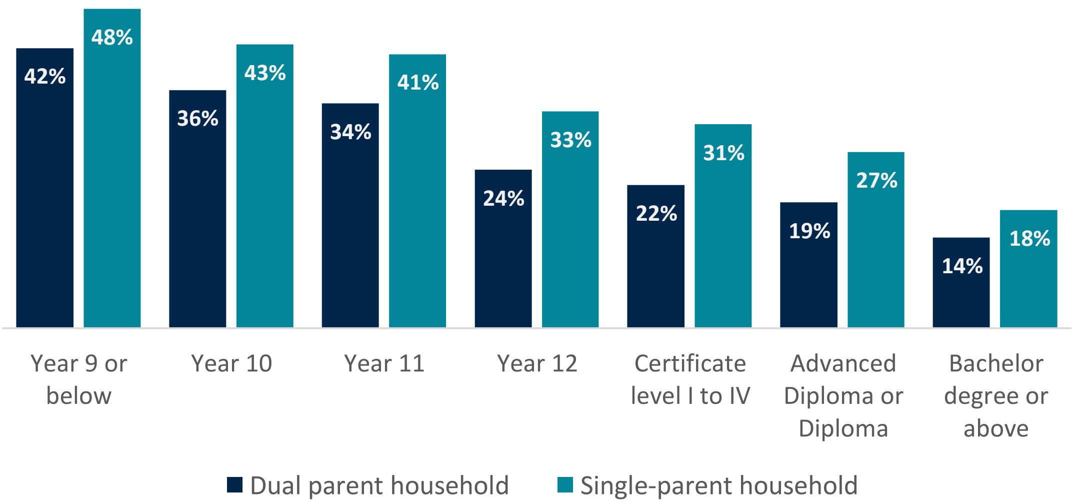
The prevalence of post-school education among parents or carers of young children in Australia is high. Only 9,756 children had at least one parent or carer with a highest educational attainment of Year 9 or below, with 2,496 children having two parents or carers with a highest educational attainment of Year 9 or below. In contrast, 64,356 children had two parents or carers with a highest educational attainment of bachelor degree or above and a total of 114,361 children had at least one parent or carer with a bachelor degree or above.

Even once a range of other factors were accounted for, our predictive modelling showed that the highest level of educational attainment of the most educated parent or carer was an important factor for predicting developmental vulnerability (see *Predictive modelling* section, Figures 1, 2b, 3b).

Single parent households

Overall, children from single parent households had higher rates of developmental vulnerability than children with two parents (29 and 18 per cent respectively), and this was true for all parental education levels (Figure 3). Developmental vulnerability of children in single parent households decreases as parent or carer educational attainment increases, ranging from 48 per cent for Year 9 or below to 18 per cent for Bachelor degree or above. Our predictive modelling identified being from a single parent household as an important factor of developmental vulnerability (see *Predictive modelling* section, Figure 1)[2].

Figure 3. Proportion (%) of children who are developmentally vulnerable on one or more domains, by highest educational attainment of most educated parent or carer, by single parent household status, 2018 AEDC cohort.



Source: Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.

Notes: This figure compares children from the 2018 cohort of the Australian Early Development Census. N =252,998. Single-parent household status was identified through child and parent(s) shared address data. Parent or carer education was identified using the AEDC (see Methodology section).

[2] The parent or carer variable in AEDC does not individually identify the parent or carer and cannot be directly linked to the parent identity used in the single parent household information.

Neighbourhood-level characteristics such as the proportion of single parent households, the proportion of unemployed adults and socio-economic status (SES) in general have also been linked to developmental outcomes (Leventhal and Brooks-Gunn 2000).

In the 2018 AEDC cohort, children from single parent households were more represented in lower household income deciles, ranging from 54 per cent in the first decile to only 9 per cent in the tenth decile. This trend is also true for lower neighbourhood SES, ranging from 45 per cent single-parent households in the lowest SES neighbourhood, to 13 per cent in the highest SES neighbourhoods. The benefit of two parents or carers compared to one could be attributed to the capacity to share the parental load and higher income due to a greater overall capacity to work.

Neighbourhood SES and household income

Household income and neighbourhood SES are positively associated with developmental outcomes (Brooks-Gunn et al. 1993; Willms 2002). Children belonging to higher SES neighbourhoods tend to have better developmental outcomes (Leventhal and Brooks-Gunn 2000).

As the SES of the neighbourhood increased, the proportion of children who were developmentally vulnerable on one or more domains generally decreased. Within each household income decile, this trend was consistent, with children from households in lower SES neighbourhoods having higher rates of developmental vulnerability (Figure 4).

The association between lower SES neighbourhoods and poorer developmental outcomes is consistent with the literature, which suggests lower SES is associated with a lack of resources for establishing a positive developmental environment (Lapointe et al. 2007; Oliver et al. 2007). The neighbourhood setting of the child has a substantial impact upon child development as it creates the social context that directly and indirectly affects child development including school readiness upon entry (Lapointe et al. 2007; Oliver et al. 2007; Venetsanou and Kambas 2010; Coulton et al. 2016).

Similarly, as the household income decile decreased the proportion of children who were developmentally vulnerable on one or more domains generally increased. This trend held across household income deciles, however it peaked at the third household income decile before slightly declining in the lowest two household income deciles (Figure 4). This trend is somewhat surprising as lower household income is associated with greater risk of developmental vulnerability (Lapointe et al. 2007). This may reflect unmeasured wealth and relative advantage in the lowest household income earners. The Survey of Income and Housing excludes the 1st and 2nd percentiles in the low income quintile definition due to the high wealth and expenditure characteristics those households exhibit, and the prevalence of income types other than employee income and government pensions and allowances (ABS 2013; ABS 2022).

Figure 4. Proportion (%) of children who are developmentally vulnerable on one or more domains, by neighbourhood socio-economic status (SES), by household income decile, 2018 AEDC cohort.

Household income decile	Neighbourhood SES 1	2	3	4	5	6	7	8	9	10	All SES
1	35%	28%	23%	24%	19%	20%	19%	18%	17%	14%	25%
2	37%	29%	25%	24%	24%	20%	19%	20%	17%	16%	28%
3	38%	29%	26%	25%	24%	22%	21%	20%	16%	18%	29%
4	37%	28%	24%	22%	20%	18%	18%	17%	16%	16%	26%
5	34%	24%	22%	19%	18%	17%	15%	15%	15%	14%	22%
6	32%	23%	21%	19%	16%	15%	15%	14%	14%	13%	20%
7	28%	20%	18%	17%	15%	15%	15%	13%	13%	13%	18%
8	26%	19%	18%	16%	16%	14%	15%	12%	14%	13%	16%
9	23%	17%	16%	15%	14%	14%	13%	13%	12%	11%	14%
10	18%	15%	14%	14%	14%	12%	12%	12%	12%	10%	12%
All incomes	34%	25%	21%	20%	18%	16%	15%	14%	13%	12%	

Source: Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.

Notes: This figure compares children from the 2018 cohort of the Australian Early Development Census. N = 266,299. Neighbourhood SES (1 - lowest SES to 10 - highest SES) based on Socio-Economic Indexes for Areas Index of Relative Socio-Economic Advantage and Disadvantage (SEIFA IRSAD). Household income deciles are based on parental income as per tax data (see Methodology section).

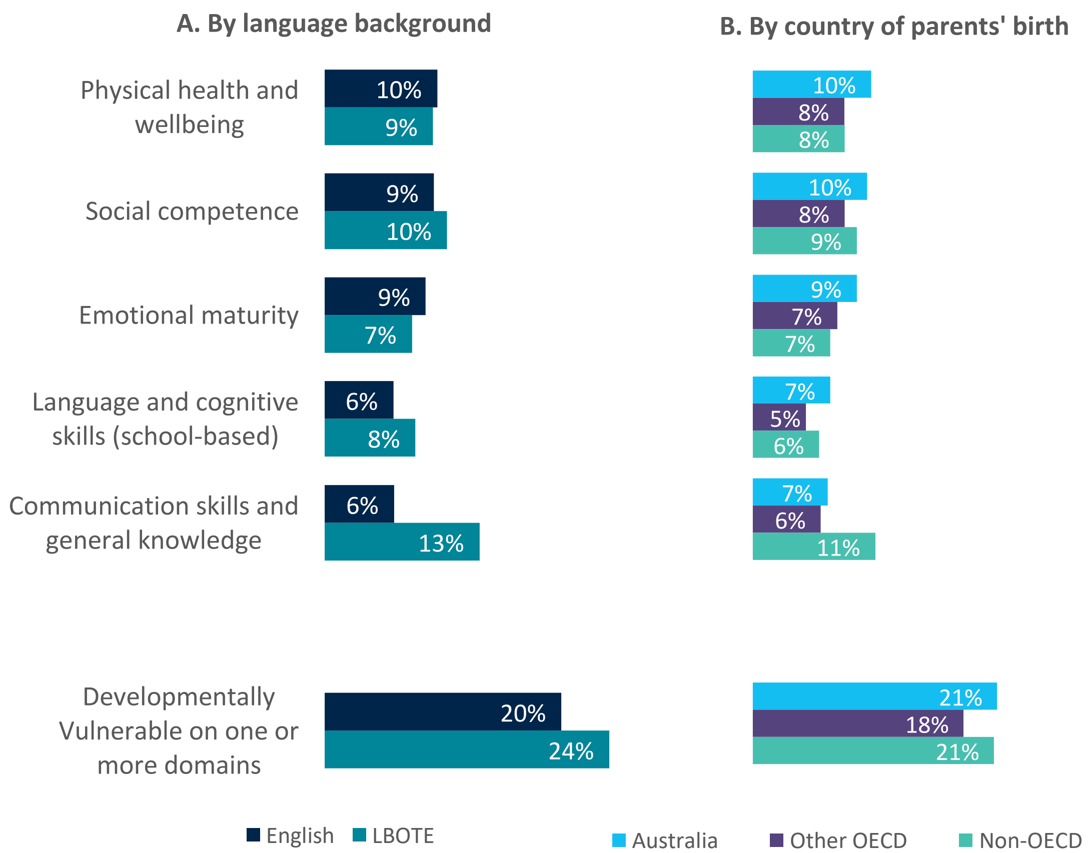
To better understand the differences between neighbourhood SES and household incomes we explored co-occurring developmental risk factors. The third-lowest household income decile contained the highest percentage of children whose most educated parent or carer had Year 10 or below as their highest educational attainment, almost double that of the first household income decile. The welfare profiles also differed in the third household income decile; there was a higher percentage of children with parents receiving rent assistance compared to the first household income decile. Additionally, children in the third household income decile had higher proportions of parents receiving carer payments, disability support pension, and unemployment payments. Both rent assistance and unemployment payments require asset tests to be eligible, potentially indicating the third decile has less wealth than all other income deciles including the first household income decile.

Our results show that family and neighbourhood socio-economic characteristics have an important association with childhood development, consistent with earlier literature on the topic – lower neighbourhood SES and lower household income are generally associated with higher rates of developmental vulnerability. Future research in this area could benefit from further unpacking these trends, including the interactions between individual risk factors and neighbourhood socio-economic attributes.

Language background and country of parent's birth

A greater proportion of children from language backgrounds other than English (LBOTE) were developmentally vulnerable on one or more domains than children from predominantly English-speaking homes in the 2018 AEDC cohort (Figure 5A) (AEDC 2019). This difference was driven by the greater proportion of children from LBOTE who were developmentally vulnerable on the communication skills and general knowledge domain. These children generally had similar rates of developmental vulnerabilities in other domains to children from English-speaking backgrounds.

Figure 5. Proportion (%) of children who are developmentally vulnerable, by domain, 2018 AEDC cohort.



Source: Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.

Notes: This figure compares children from the 2018 cohort of the Australian Early Development Census. N = 272,626. LBOTE: Language Background Other Than English based on AEDC data; OECD: Organisation for Economic Cooperation and Development based on combined demographic data.

By controlling for a wide range of events and circumstances surrounding the child, including ECEC attendance, single parent status, health and welfare receipt, we found that being from LBOTE was associated with a statistically significant increase in the likelihood of being developmentally vulnerable on at least one domain for both boys and girls (see *Predictive modelling section*, Figure 2a and 3a).

We also compared developmental vulnerability rates for children with parents born in Australia, in other OECD countries, and in non-OECD countries. OECD and non-OECD countries were selected for comparison because OECD countries, including Australia, have much higher Gross Domestic Product per capita on average than non-OECD countries[3]. Children with at least one parent born overseas in an OECD country had lower rates of developmental vulnerability on one or more domains (18 per cent) than either children with two Australian-born parents or children with a parent born in a non-OECD country (both 21 per cent; Figure 5B). This was driven by lower average developmental vulnerability rates in the communication skills and general knowledge, social competence, and language and cognitive skills (school-based) domains (Figure 5B). Our country of parents' birth data is sourced primarily from the 2016 Census of Population and Housing, therefore recent migrants' country of birth may be underrepresented (see *Methodology* section).

Previous research (Washbrook et al. 2012; Queensland Health 2019) suggests there is substantial variation in developmental outcomes of children of migrants relative to children of Australian-born parents, depending on their cultural and ethnic background as well as the developmental domain in question. Children from LBOTE with at least one parent from a non-OECD country have a great deal of overlap in our cohort. Future analysis could consider the combined effects of country of birth and language background on developmental vulnerability.

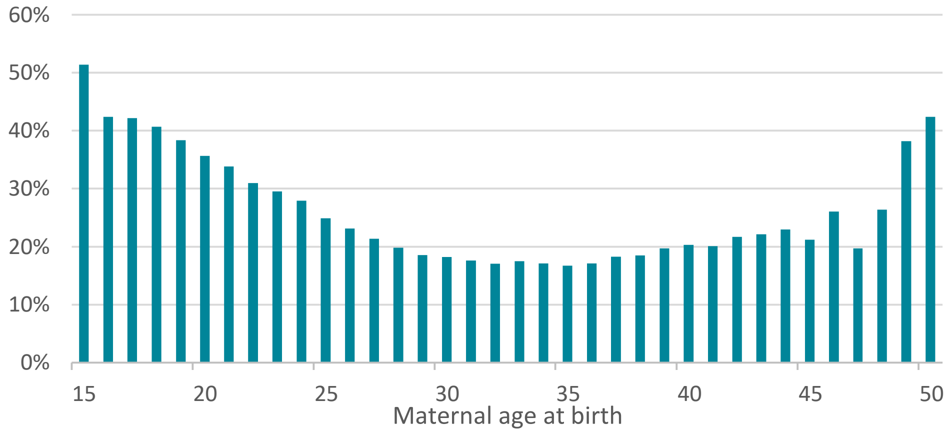
Maternal age at birth

The association between maternal age at birth and developmental vulnerability on one or more domains in the 2018 AEDC is shown in Figure 6. Rates of developmental vulnerability were highest among children of mothers under 20 years old, with 51 per cent of children of mothers aged 15 years or younger being developmentally vulnerable in one or more domains. The rate of developmental vulnerability was at its lowest point, between 16.7 and 19.8 per cent, among children born when their mothers were aged between 28 and 39 years. Rates of developmental vulnerability increased slightly among children of older mothers (Figure 6).

These findings are similar to results of previous research into maternal age at birth on child development (Falster et al. 2018; Hanly et al. 2020). However, once we adjusted the data for a range of other child and family circumstances, we found that maternal age at birth was less important in explaining developmental vulnerability than most other factors investigated, including parental employment, neighbourhood SES and child and parental mental health issues (see *Predictive modelling* section). This is consistent with other research that suggests maternal age at birth performs poorly as a sole predictor of child developmental vulnerability (Chittleborough et al. 2011).

[3] World Bank data on 2019 GDP per capita in US Dollars (World Bank 2021).

Figure 6. Proportion (%) of children who are developmentally vulnerable on one or more domains (DV1), by maternal age at birth, 2018 AEDC cohort.



Source: Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.

Notes: This figure compares children from the 2018 cohort of the Australian Early Development Census. N = 268,995.

Parental mental and physical health

Chronic physical and mental ill-health in children and their parents have been associated with poorer development outcomes in children (Bell et al. 2016; Razaz et al. 2016; Comaskey et al. 2017; Bell et al. 2019; Laurens et al. 2019).

To study the effect of physical and mental health on developmental vulnerability in our cohort, we developed two health measures using government-supported health service access as a proxy for experiencing mental ill-health and chronic physical ill-health. Accessing Medicare Benefits Schedule (MBS) items, for example psychologist visits, or Pharmaceutical Benefits Scheme (PBS) prescriptions, for example anti-depressants, were used to indicate mental ill-health. Accessing MBS items related to chronic physical ill-health, for example Chronic Disease Management, were used to indicate chronic physical ill-health (see *Methodology* section).

The analysis focussed on children with developmental data in the AEDC. The AEDC excludes domain and developmental vulnerability scores for children with special needs because of the already identified substantial developmental needs of this group (Social Research Centre 2019). Children with special needs (without developmental scores) have a higher prevalence of mental and chronic ill-health compared to children without special needs[4]. As a result, the developmental risks associated with ill-health are likely to be understated.

Most of the analysis on ill-health is performed for girls and boys separately for two reasons. First, boys were more likely to be developmentally vulnerable than girls on one or more domains (DV1 rates for boys and girls are 27.0% and 14.6% respectively), and on each individual AEDC domain (AIHW 2020). The difference in development is thought to reflect principally biological dispositions, although different parenting practices and expectations for boys and girls also might have some influence

[4] Children with special needs (as identified by the AEDC): mental ill-health 46%, chronic physical ill-health 57%, N = 13,593. Children without special needs: mental ill-health 8%, chronic physical ill-health 12%, N = 279,648.

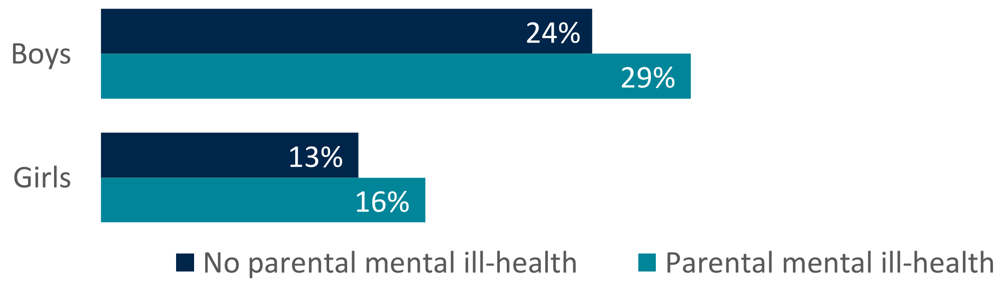
(AIHW 2015). Furthermore, our results show the prevalence of ill-health and the impact of parental ill-health are quite different for boys and girls.

Parents' mental ill-health have also been associated with poor developmental outcomes in children (Comaskey et al. 2017; Bell et al. 2019). For the 2018 cohort, we found that over half of Australian children had at least one parent experiencing mental ill-health between the year prior to their birth and 2018. Past research has been focused on maternal and perinatal issues (Field 2011; Raposa et al. 2014; Comaskey et al. 2017).

We found children who had a parent experiencing mental ill-health had higher rates of developmental vulnerability on one or more domains (Figure 7).

This trend was stronger for children whose mother had mental ill-health (boys 30 per cent; girls 17 per cent) over the father (boys 27 per cent; girls 15 per cent). This finding is consistent with several studies that indicate maternal health conditions have a higher impact on child development (Field 2011; Comaskey et al. 2017).

Figure 7. Proportion (%) of children who are developmentally vulnerable on one or more domains, by parental mental ill-health status, by gender, 2018 AEDC cohort.



Source: Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.

Notes: This figure compares children from the 2018 cohort of the Australian Early Development Census. Boys (N = 137,273); Girls (N = 134,861). Parental ill-health has been identified using MBS and PBS data from birth to 2018 (see Methodology section).

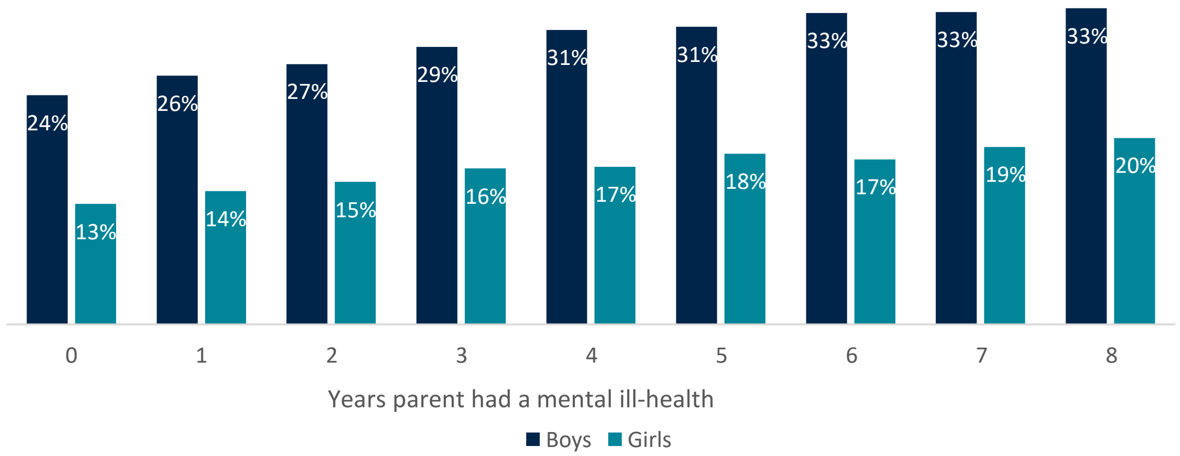
Children were much more likely to access services for mental ill-health if either of their parents accessed mental health services compared to children that had no parents with mental ill-health (10.7 compared to 3.4 per cent). Intergenerational mental ill-health is likely caused by both genetic and environmental factors (Comaskey et al. 2017). Increased parental awareness of mental ill-health may contribute to higher rates of children accessing mental health services. In 87 per cent of cases where both the child and parent accessed mental health services, the parents' mental health service access preceded the child's access.

Longitudinal mental ill-health

Past research has explored the impact of severity, timing and chronicity of parental mental ill-health on childhood development, and has shown that chronic or recurring mental ill-health are particularly detrimental (Brennan et al. 2000; Comaskey et al. 2017).

We studied the history of parental mental health service access from one year prior to childbirth to 2018 (Figure 8). As the number of years a parent experienced mental ill-health increased, the risk of developmental vulnerability trended upwards.

Figure 8. Proportion (%) of children who are developmentally vulnerable on one or more domains, by number of years of parental mental ill-health and gender, 2018 AEDC cohort.



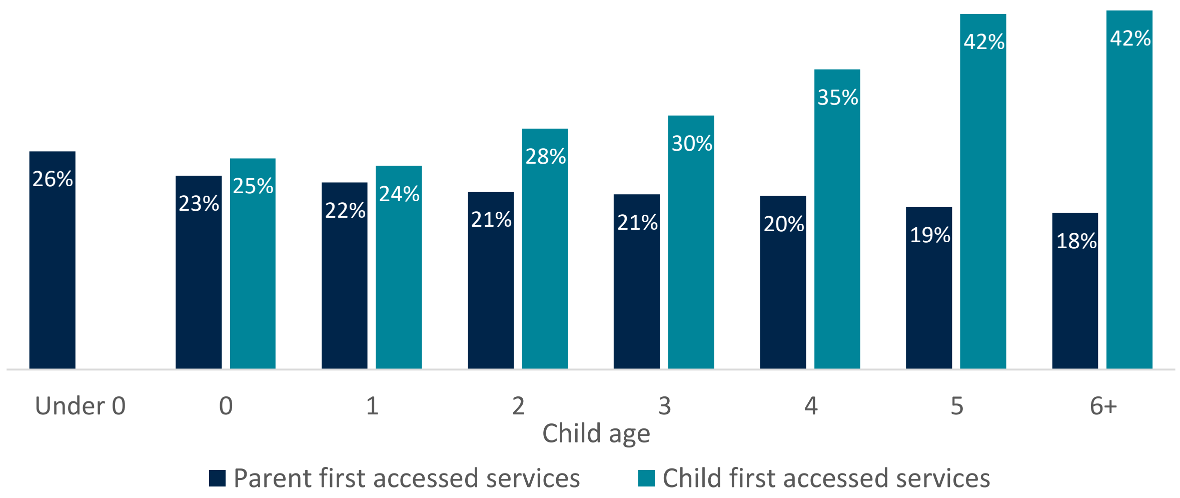
Source: Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.

Notes: This figure compares children from the 2018 cohort of the Australian Early Development Census. Boys (N = 137,273); Girls (N=134,861). Parental ill-health was identified using MBS and PBS data from one year prior to birth to 2018 (see Methodology section).

The timing of mental ill-health and relationship with development is also an area of interest. Past work indicates that early exposure to parental mental ill-health, particularly during perinatal periods, is a risk factor for childhood development outcomes (Sohr-Preston and Scaramella 2006; Field 2011; Raposa et al. 2014). We analysed developmental vulnerability in relation to a child's age when the parents first accessed services during the child's lifetime and child mental ill-health. Note that mental ill-health is often chronic or episodic, so people may dip in and out of service use over a long period of time. Barriers to accessing mental health services also make it unlikely that access to a service is instantaneous with the onset of symptoms.

Parental early access of mental health services was related to higher rates of developmental vulnerability (Figure 9). This could reflect a child's long-term exposure to an unwell parent, or exposure during a critical stage of child development (Raposa et al. 2014; Comaskey et al. 2017). Conversely, for children, later access of mental health services was related to higher rates of developmental vulnerability (Figure 9). Early diagnosis and treatment of mental ill-health may contribute to improved coping mechanisms and better developmental outcomes by the time of the AEDC. However, it could reflect that different types of mental ill-health tend to be diagnosed at different ages.

Figure 9. Proportion (%) of children who are developmentally vulnerable on one or more domains, by child age when parents/children first accessed services during the child's lifetime, 2018 AEDC cohort.



Source: Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.

Note: This figure compares children from the 2018 cohort of the Australian Early Development Census. Only children with child mental ill-health (N = 20,941) or parental mental ill-health (N = 160,062) are shown. Parent ill-health was identified from one year prior to birth to 2018. Child health conditions were identified using MBS and PBS data from birth to 2018 (see Methodology section).

Family and social factors on mental health

Our predictive modelling considered a wide variety of family and individual factors and found child health factors to have high importance for boys and girls in predicting developmental vulnerability. Meanwhile, parental health factors were important in predicting developmental vulnerability for girls (see *Methodology* and *Predictive modelling* sections).

Previous research has indicated that family and social factors such as dual parenthood can offset the risk of poor health outcomes (Reupert et al. 2013). We found that children in single parent households were more likely to access mental health services than children in dual parent families at 10.4 per cent and 6.7 per cent, respectively. The rate of developmental vulnerability of children with mental ill-health decreased from 42 per cent in single parent families to 33 per cent in partnered parent families.

The incidence of child mental ill-health ranges from 7 to 9 per cent, with a slightly higher incidence associated with lower socio-economic status. Child mental ill-health was less prevalent in children from a language background other than English (4 per cent) compared to children from English language backgrounds (9 per cent). It is not clear if this reflects the difference in prevalence of mental ill-health in the two populations, a difference in the rate of mental health service access in these two populations, or a combination of both.

Future research in this area will benefit from unpacking the interactions between other factors and mental health in greater depth, as well as using other health data such as hospitalisations. There is also the need to specifically examine the impacts of early infant and child maltreatment and increased risk of developmental vulnerability, given the direct correlation established between child maltreatment and increased prevalence of mental health disorders (Scott et al. 2023).

References

ABS (Australian Bureau of Statistics) (2013) 'Household economic wellbeing fact sheets', *6523.0 – Household Income and Income Distribution, Australia, 2011–12,* ABS, accessed 3 September 2024.

ABS (2022) *Household Income and Wealth, Australia, 2019–20*, ABS, accessed 3 September 2024

AEDC (Australian Early Development Census) (2019) *About the AEDC domains*, AEDC website, accessed 18 February 2021.

AIHW (Australian Institute of Health and Welfare) (2015) *Literature review of the impact of early childhood education and care on learning and development*, catalogue number CWS 53, AIHW, Australian Government, accessed 3 September 2024.

AIHW (2020) *Australia's children*, catalogue number CWS 69, AIHW, Australian Government, accessed 26 February 2021.

Bell MF, Bayliss DM, Glauert R, Harrison A and Ohan JL (2016) 'Chronic illness and developmental vulnerability at school entry', *Pediatrics,* 137(5):e20152475, doi:10.1542/peds.2015-2475.

Bell MF, Bayliss DM, Glauert R, Harrison A and Ohan JL (2019) 'Children of parents who have been hospitalised with psychiatric disorders are at risk of poor school readiness', *Epidemiology and Psychiatric Sciences,* 28(5):508–520, doi:10.1017/S2045796018000148.

Brennan PA, Hammen C, Andersen MJ, Bor W, Najman JM and Williams GM (2000) 'Chronicity, severity, and timing of maternal depressive symptoms: relationships with child outcomes at age 5', *Developmental Psychology,* 36(6):759–766, doi:10.1037//0012-1649.36.6.759.

Brooks-Gunn J, Duncan G, Klebanov PK and Sealand N (1993) 'Do neighborhoods influence child and adolescent development?', *American Journal of Sociology*, 99(2):353–395, doi:10.1086/230268.

Chittleborough CR, Lawlor DA and Lynch JW (2011) 'Young maternal age and poor child development: predictive validity from a birth cohort', *Pediatrics,* 127(6):1436–1444, doi:10.1542/peds.2010-3222.

Comaskey B, Roos NP, Brownell M, Enns MW, Chateau D, Ruth CA and Ekuma O (2017) 'Maternal depression and anxiety disorders (MDAD) and child development: a Manitoba population-based study', *PLoS One,* 12(5):e0177065, doi:10.1371/journal.pone.0177065.

Comuk-Balci N, Bayoglu B, Tekindal A, Kerem-Gunel M and Anlar B (2016) 'Screening preschool children for fine motor skills: environmental influence', *Journal of Physical Therapy Science*, 28(3):1026–1031, doi:10.1589/jpts.28.1026.

Coulton CJ, Richter F, Kim S-J, Fischer R and Cho Y (2016) 'Temporal effects of distressed housing on early childhood risk factors and kindergarten readiness', *Children and Youth Services Review*, 68:59–72, doi:10.1016/j.childyouth.2016.06.017.

DET (Department of Education and Training) (2019) *International Standard Classification of Education 2011 (ISCED 2011) to Australian Standard Classification of Education (ASCED) Concordance*, accessed January 2024.

Falster K, Hanly M, Banks E, Lynch J, Chambers G, Brownell M, Eades S and Jorm L (2018) 'Maternal age and offspring developmental vulnerability at age five: a population-based cohort study of Australian children', *PLoS Medicine,* 15(4):e1002558, doi:10.1371/journal.pmed.1002558.

Field T (2011) 'Prenatal depression effects on early development: a review', *Infant Behavior and Development,* 34(1):1–14, doi:10.1016/j.infbeh.2010.09.008.

Hanly M, Falster K, Banks E, Lynch J, Chambers GM, Brownell M, Dillon A, Eades S and Jorm L (2020) 'Role of maternal age at birth in child development among Indigenous and non-Indigenous Australian children in their first school year: a population-based cohort study', *The Lancet Child & Adolescent Health,* 4(1):46–57, doi:10.1016/s2352-4642(19)30334-7.

Kohl HW and Hobbs KE (1998) 'Development of physical activity behaviours among children and adolescents', *Pediatrics*, 101(Supplement 2):549–554, doi:10.1542/peds.101.S2.549.

Lapointe VR, Ford L and Zumbo BD (2007) 'Examining the relationship between neighborhood environment and school readiness for kindergarten children', *Early Education and Development*, 18(3):473–495, doi:10.1080/10409280701610846.

Laurens KR, Green MJ, Dean K, Tzoumakis S, Harris F, Islam F, Kariuki M, Essery CM, Schofield JM and Carr VJ (2019) 'Chronic physical health conditions, mental health, and sources of support in a longitudinal Australian child population cohort', *Journal of Pediatric Psychology,* 44(9):1083–1096, doi:10.1093/jpepsy/jsz048.

Leventhal T and Brooks-Gunn J (2000) 'The neighbourhoods they live in: the effects of neighbourhood residence on child and adolescent outcomes', *Psychological Bulletin*, 126(2):309–337, doi:10.1037/0033-2909.126.2.309.

Noble KG, Houston SM, Brito NH, Bartsch H, Kan E, Kuperman JM, Akshoomoff N, Amaral DG, Bloss CS, Libiger O, Schork NJ, Murray SS, Casey BJ, Chang L, Ernst TM, Frazier JA, Gruen JR, Kennedy DN, Van Zijl P, Mostofsky S, Kaufmann WE, Kenet T, Dale AM, Jernigan TL and Sowell ER (2015) 'Family income, parental education and brain structure in children and adolescents', *Nature Neuroscience*, 18(5):773–778, doi:10.1038/nn.3983.

Oliver LN, Dunn JR, Kohen DE and Hertzman C (2007) 'Do neighbourhoods influence the readiness to learn of kindergarten children in Vancouver? A multilevel analysis of neighbourhood effects', *Environment and Planning A*, 39(4), 848–868, doi:10.1068/a37126.

Queensland Health (2019) *Culturally and linguistically diverse children and their families – implications for paediatric and child development services in Queensland*, Queensland Health, Queensland Government, accessed 8 January 2021.

Raposa E, Hammen C, Brennan P and Najman J (2014) ‘The long-term effects of maternal depression: early childhood physical health as a pathway to offspring depression’, *Journal of Adolescent Health,* 54(1):88–93, doi:10.1016/j.jadohealth.2013.07.038.

Razaz N, Joseph KS, Boyce WT, Guhn M, Forer B, Carruthers R, Marrie RA and Tremlett H (2016) ‘Children of chronically ill parents: relationship between parental multiple sclerosis and childhood developmental health’, *Multiple Sclerosis,* 22(11):1452–1462, doi:10.1177/1352458515621624.

Reupert AE, Maybery DJ and Kowalenko NM (2013) ‘Children whose parents have a mental illness: prevalence, need and treatment’, *The Medical Journal of Australia,* 199(3):S7–S9, doi:10.5694/mja11.11200.

Scott JG, Malacova E, Mathews B, Haslam DM, Pacella R, Higgins DJ, Meinck F, Dunne MP, Finkelhor D, Erskine HE, Lawrence DM and Thomas HJ (2023) The association between child maltreatment and mental disorders in the Australian Child Maltreatment Study. *Medical Journal of Australia*, 218: S26–S33, doi:10.5694/mja2.51870.

Social Research Centre (2019) *2018 AEDC Data Collection Technical Report*, report to Australian Government Department of Education, Social Research Centre, accessed January 2024

Sohr-Preston SL and Scaramella LV (2006) ‘Implications of timing of maternal depressive symptoms for early cognitive and language development’, *Clinical Child and Family Psychology Review,* 9(1):65–83, doi:10.1007/S10567-006-0004-2.

Venetsanou, F and Kambas A (2010) ‘Environmental factors affecting preschoolers’ motor development’, *Early Childhood Education Journal*, 37(4):319–327, doi:10.1007/s10643-009-0350-z.

Washbrook E, Waldfogel J, Bradbury B, Corak M and Ghanghro AA (2012) ‘The development of young children of immigrants in Australia, Canada, the United Kingdom, and the United States’, *Child Development*, 83(5):1591–1607, doi:10.1111/j.1467-8624.2012.01796.x.

Willms JD (2002) ‘Socioeconomic gradients for childhood vulnerability’, in Willms JD (ed) *Vulnerable children: findings from Canada’s national longitudinal survey of children and youth,* University of Alberta Press, Edmonton.

World Bank (2021) *GDP per capita, (current US$)* [data set], World Bank website, accessed 8 February 2021.

The First Five Years: What makes a difference?

2.2 Predictive modelling of developmental vulnerability in children

Key findings

- Many factors are associated with developmental vulnerability in children. The most significant factors identified by our predictive models are parental employment, parental education, neighbourhood SES, preschool attendance, and child chronic ill-health.
- These factors affect boys and girls differently. For example, child mental ill-health is associated with higher rates of child developmental vulnerability in boys.
- As well as the direct association between child care attendance and children's developmental vulnerability, child care is an enabler for other factors that were important such as parental employment.

In the previous sections, descriptive analysis of the socio-economic factors and physical and mental health factors affecting children's developmental vulnerability were presented, without considering cross-combinations with other factors that may impact developmental vulnerability.

This section explores which factors, out of all the available variables, predict developmental vulnerability (DV1) and estimates the order and magnitude of their importance. The predictive modelling mines the data for patterns that have predictive power, without drawing on theories of change or subject matter knowledge to assess the plausibility and direction of causal relationships between the factors and developmental outcomes. As such, it does not differentiate between factors that may have some causal link with child development, and those that are only correlated (including with unobserved factors). A combination of regression and machine learning techniques are used to rank the variables based on their predictive power (see details in the *Methodology* section).

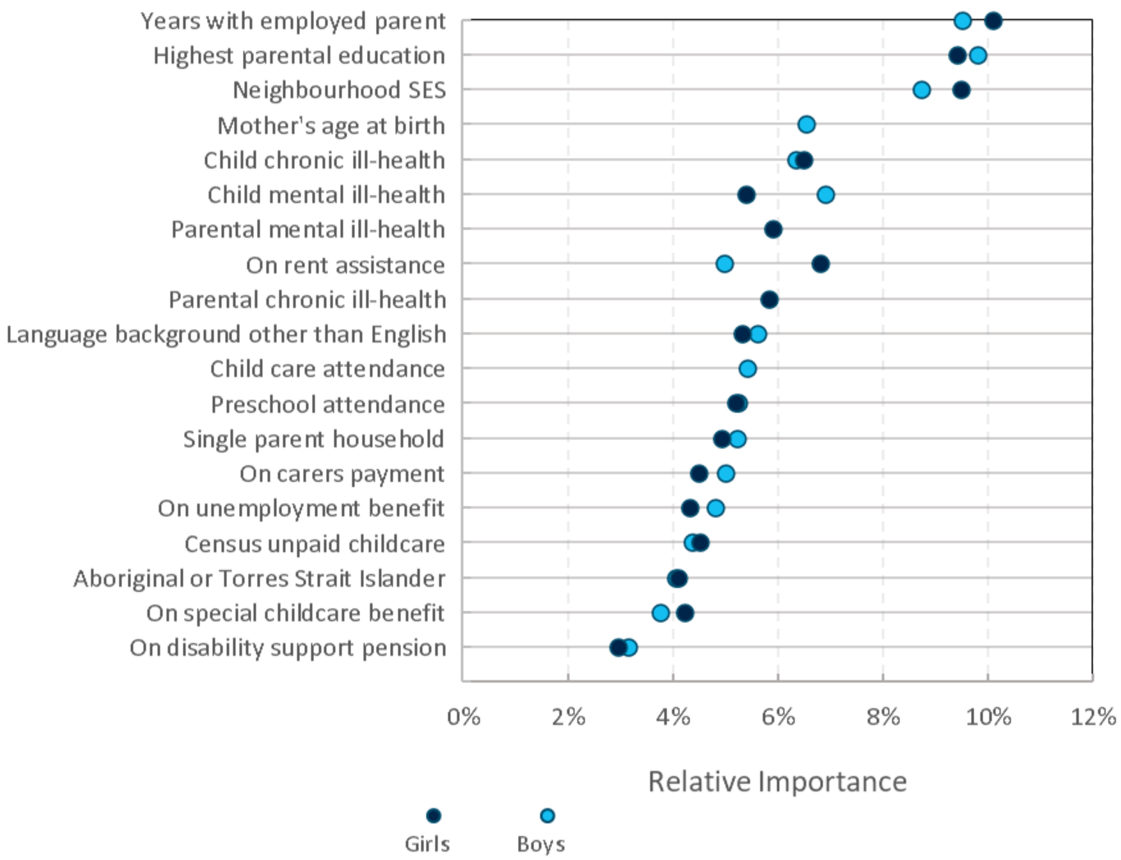
Which factors are important in predicting developmental vulnerability?

Figure 1 displays the variable importance rankings for predicting developmental vulnerability. For each variable, the variable importance shows the percentage change in accuracy of predicting the children's developmental outcome when the variable is removed. The higher the percentage change, the higher the variable importance. Developmental outcomes differed for boys and girls, so separate models were constructed for each gender. The variable importance was calculated after subsetting the factors via stepwise regression, so some factors do not appear in this figure, and boys and girls may have different selected factors. For example, *Maternal age at birth* is present only for boys and *Parental mental health issue* is present only for girls.

Years with employed parent, *Highest parental education* and *Neighbourhood SES* factors had the strongest predictive power for both genders. Parental income was not selected by the stepwise logistic regression as a predictor of developmental vulnerability. This is likely because it is highly correlated with *highest parental education* and *years with an employed parent*.

In general, the mental and chronic health of the child as well as established socio-economic disadvantages and parental health were important for predicting developmental vulnerability.

Figure 1. Variable importance rankings for selected factors on developmental vulnerability, by gender, AEDC 2018 cohort.



Source: Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.

Notes: This figure uses data from the 2018 cohort of the Australian Early Development Census. DV1: Developmentally vulnerable on one or more domains. Boys (N = 90,577); Girls (N = 88,837). Permutation importance was calculated and scaled such that the importance values sum to 100 per cent. F1 scores[1] for boys: 0.38 and for girls: 0.23.

There are some differences between the factor importance lists for boys and girls. For example, *Rent assistance* appears more important for girls, and *Child mental ill-health* is more important for boys.

How do these factors affect developmental vulnerability?

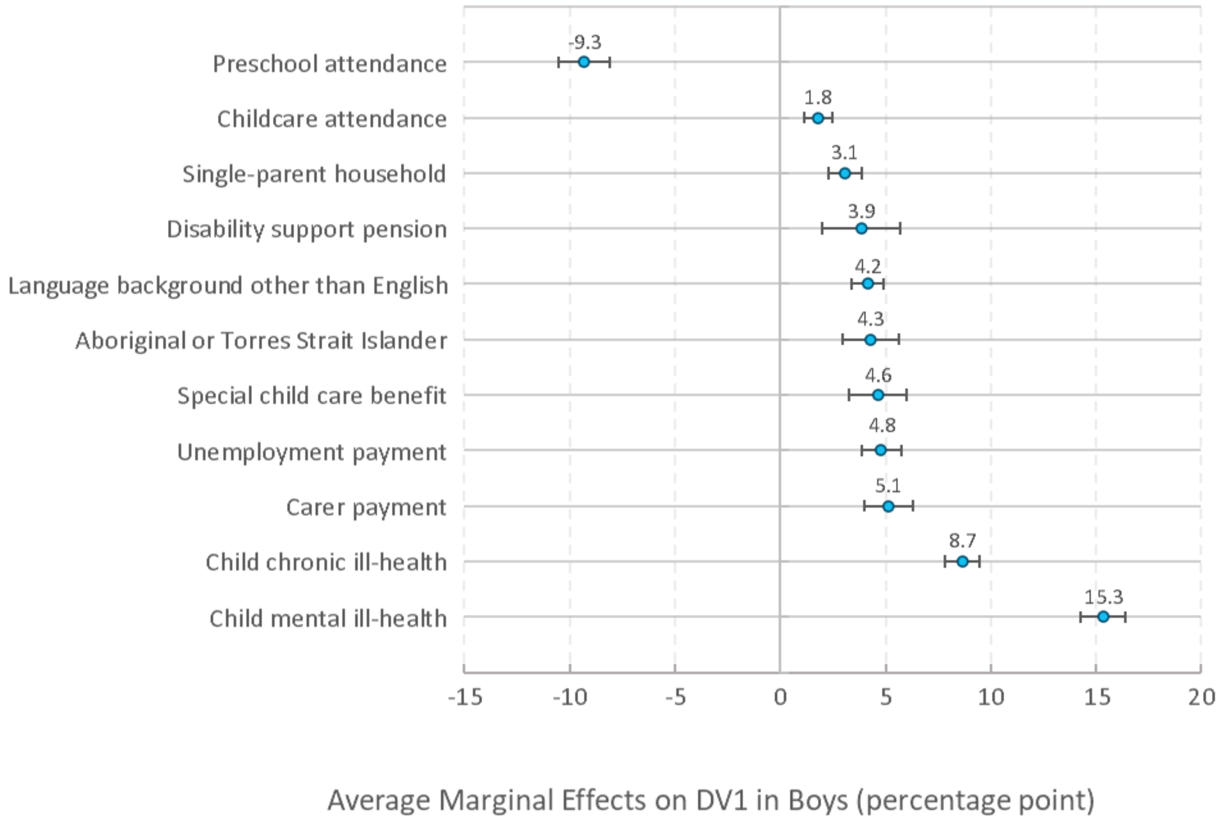
We calculated average marginal effects (AMEs) for the selected factors from the stepwise logistic regression. The AME gives an indication of a factor's impact on the likelihood of a child being developmentally vulnerable, compared to the reference group. For example, for the preschool attendance variable, the AME gives the average change in the rate of DV1 for the attended group comparing to the non-attended group. A positive AME indicates a factor is positively correlated with

[1] F1 score is a measure of the performance of the model prediction, ranging from 0 to 1. A score of 1 means that the model can predict the outcome perfectly. In our models, the low F1 score is an indication of the many factors affecting child development, some of which cannot be measured.

developmental vulnerability while a negative AME indicates the factor is negatively correlated with developmental vulnerability.

We calculated AMEs for both binary and multi-level factors. A binary factor only takes two values, e.g., yes or no (with no as the baseline level). For example, the *child care* factor is a measure of whether a child ever attended childcare. It does not reflect other aspects, such as the number of hours, the quality or type of child care they attended. A multi-level factor has three or more discrete values, either with or without an order, for example *highest parental education*. AMEs for multi-level factors are compared to a baseline level listed under the factor name in the figures, e.g., the baseline level for *highest parental education* is *Year 9 or below*.

Figure 2a. Average marginal effects for selected binary factors in boys' developmental vulnerability on one or more domains, 2018 AEDC cohort.



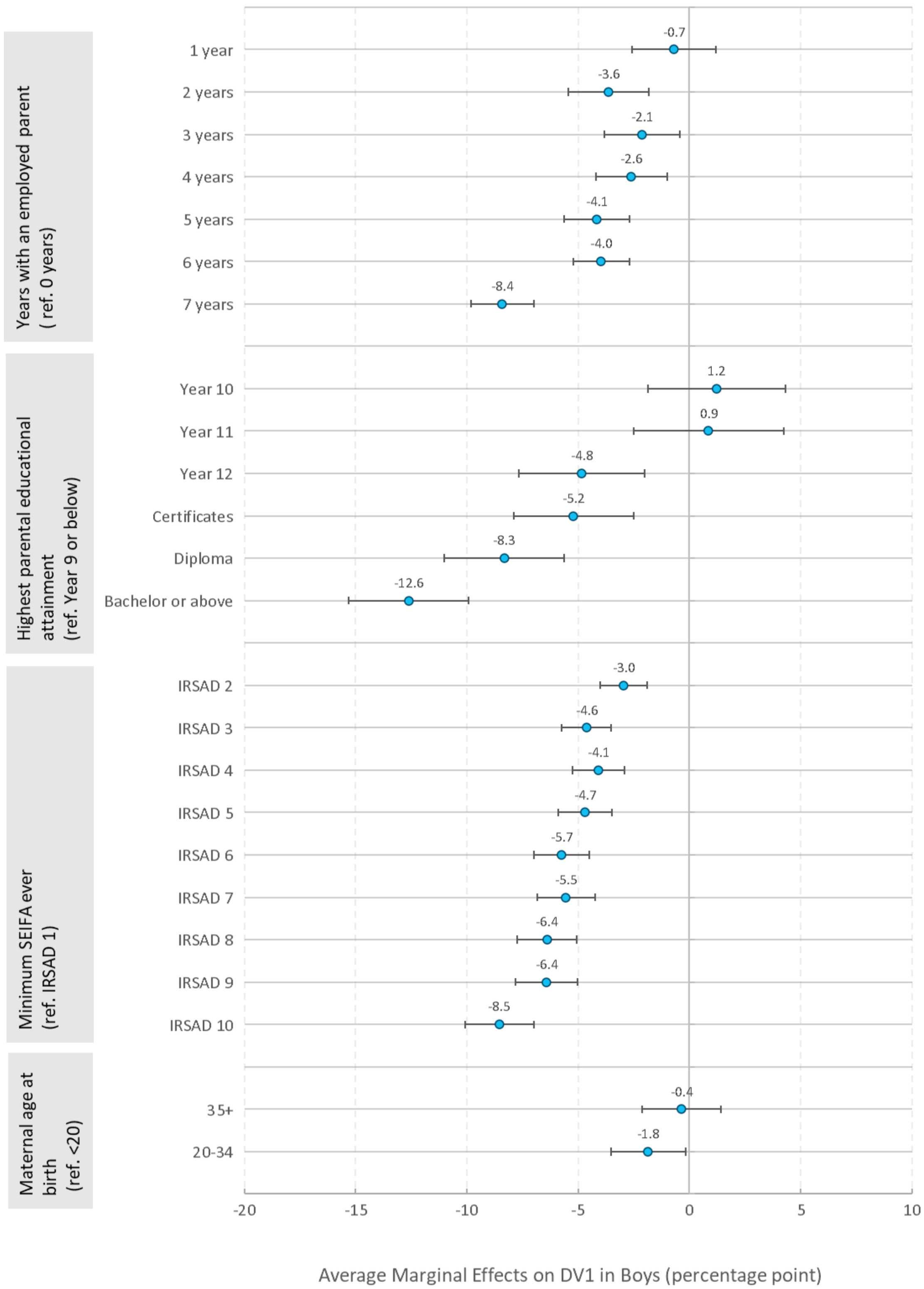
Source: Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.

Notes: This figure uses data from the 2018 cohort of the Australian Early Development Census. DV1: Developmentally vulnerable on one or more domains. Error bars are 95 per cent confidence intervals. N = 90,577. For technique and factor descriptions, see the *Methodology* section.

Similarly to the factor importance analyses, mental ill-health appears to have a strong correlation with boys' developmental vulnerability, increasing the likelihood of developmental vulnerability by 15 percentage points on average (Figure 2a). At the other end of the spectrum, preschool is associated with a reduction in developmental vulnerability. It may be that having a year of school preparation prior to the AEDC is helping these children's scores. Though preschool attendance has a large AME, in Fig.1, preschool attendance does not appear to have a large effect in predicting the children's developmental vulnerability, since 95 per cent of the 2018 AEDC cohort attended preschool.

Figure 2b shows the results for boys' multi-level factors. Highest parental educational attainment of most educated parent, higher Index of Relative Socio-economic Advantage and Disadvantage (IRSAD) and more years with employed parents are negatively correlated with boys' developmental vulnerability. Bachelor degree or above *highest parental educational attainment* appears to have the strongest of these effects (-13 percentage points) when compared to the parental education baseline of below Year 10.

Figure 2b. Average marginal effects of multi-level factors in boys' developmental vulnerability on one or more domains, 2018 AEDC cohort.

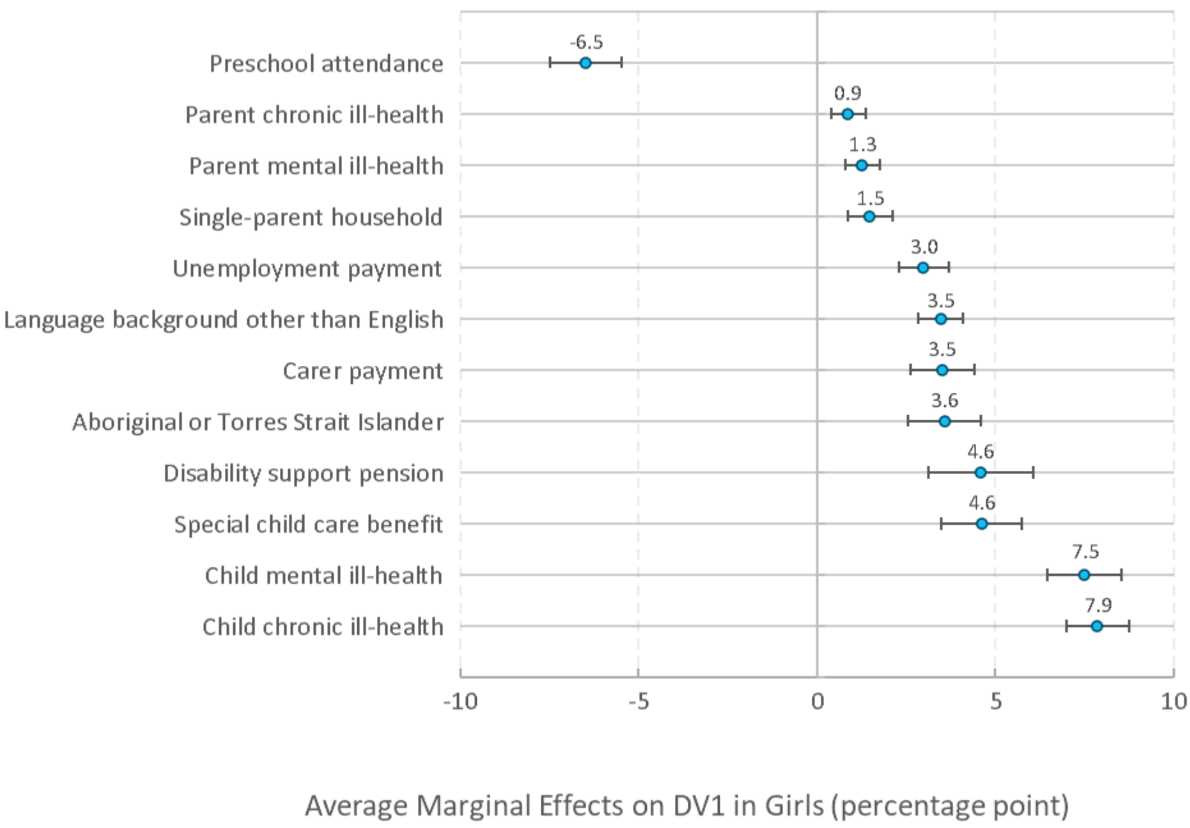


Source: Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.

Notes: This figure uses data from the 2018 cohort of the Australian Early Development Census. DV1: Developmentally vulnerable on one or more domains. Error bars are 95 per cent confidence intervals. N= 90,577. SEIFA: Socio Economic Indexes for Areas. IRSAD: Index of Relative Socio-economic Advantage and Disadvantage (ABS 2018). IRSAD summarises information about the economic and social conditions of people and households within an area, including both relative advantage and disadvantage measures.

Figure 3a shows the AMEs for girls' binary factors. Girls' developmental vulnerability seems to be influenced less by the binary factors compared to boys' developmental vulnerability. Boys' binary AMEs range between -9.3 and 15 percentage points whereas girls' range between -6.5 and 7.9 percentage points. Most notably, the AME for mental ill-health for girls is half that of boys.

Figure 3a. Average marginal effects for selected binary factors in girls' developmental vulnerability on one or more domains, 2018 AEDC cohort.

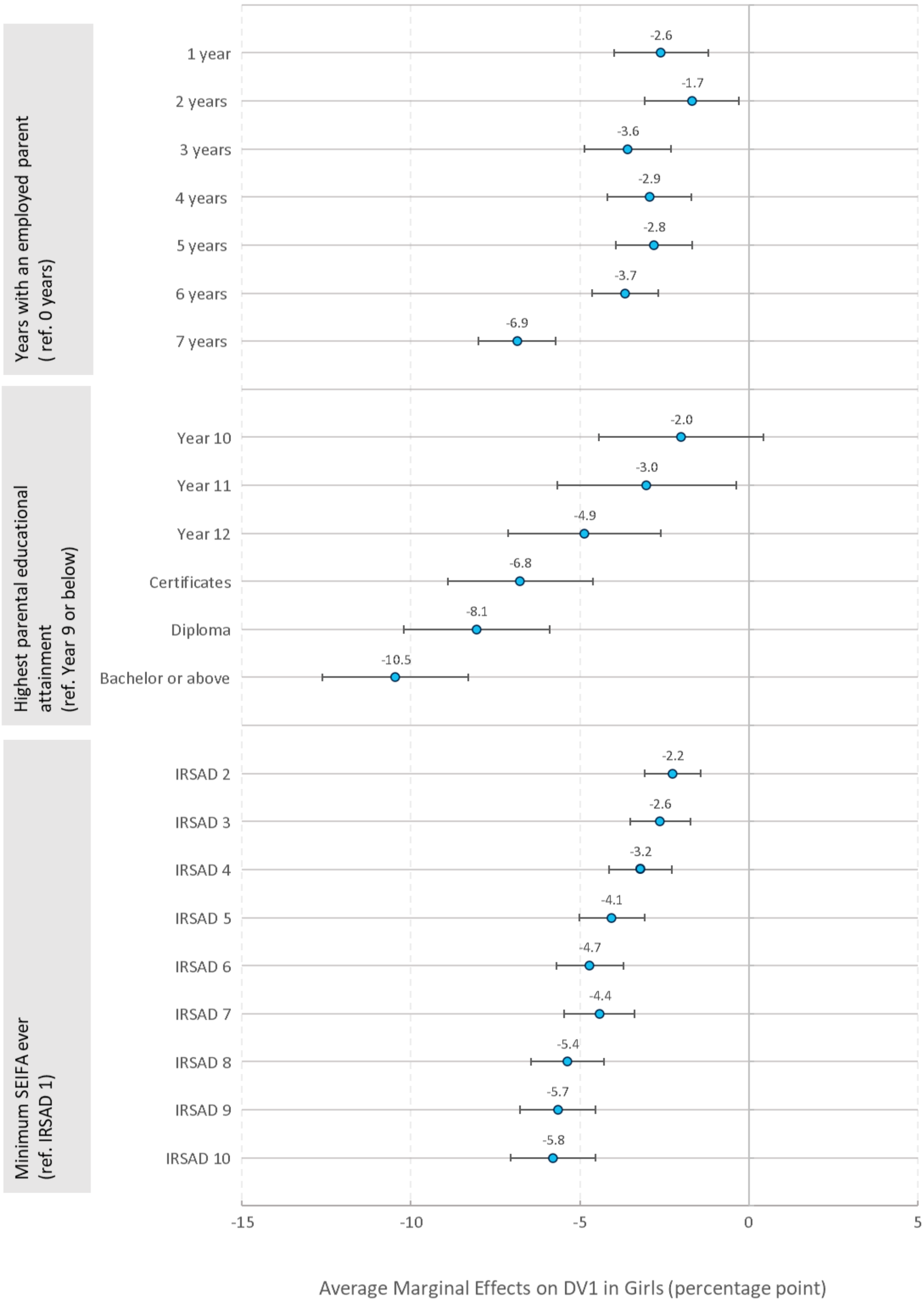


Source: Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.

Notes: This figure uses data from the 2018 cohort of the Australian Early Development Census. DV1: Developmentally vulnerable on one or more domains. Error bars are 95 per cent confidence intervals. N = 88,837. For technique and factor descriptions, see the Methodology section.

Figure 3b shows the AMEs for multi-level factors for girls. Similarly to the binary indicators, the magnitude of these AMEs for girls are smaller than for boys. For example, the AME for highest parental educational attainment of Bachelor degree or above for girls is -10 percentage points compared to -13 percentage points for boys. A similar trend can be seen for children in IRSAD 10 for neighbourhood SES, -5.8 percentage points for girls and -8.5 percentage points for boys. This suggests that boys' development vulnerability may be more sensitive to these observed external factors when compared to girls.

Figure 3b. Average marginal effects of multi-level factors in girls' developmental vulnerability on one or more domains, 2018 AEDC cohort.



Source: Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.

Notes: This figure uses data from the 2018 cohort of the Australian Early Development Census. DV1: Developmentally vulnerable on one or more domains. Error bars are 95 per cent confidence intervals. N = 88,837. SEIFA: Socio Economic Indexes for Areas. IRSAD: Index of Relative Socio-economic advantage and disadvantage (ABS 2018). For technique and factor explanations, see the Methodology section.

References

ABS (Australian Bureau of Statistics) (2018) *2033.0.55.001 — Census of Population and Housing: Socio-Economic Indexes for Areas (SEIFA), Australia, 2016*, ABS, accessed 5 January 2021.

AEDC (Australian Early Development Census) (2019) *About the AEDC domains*, AEDC website, accessed 18 February 2021.

Australian Government
Department of Education

The First Five Years: What makes a difference?

3.1 Associations between early childhood education and care attendance and child development

Key findings

We analysed the associations between early childhood education and care attendance and developmental vulnerability, including the influence of other characteristics.

- For children from Aboriginal and Torres Strait Islander families, language other than English and single parent households, attending formal child care resulted in higher rates of being developmentally on track on all 5 domains.
- 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 developmental domains showed a different pattern.
- Rates of formal child care attendance decreased with distance from metropolitan areas.
- Rates of formal child care attendance increased for higher levels of parental education.
- Children who attended preschool had much lower rates of developmental vulnerability.

Child care attendance

Data and definitions

Child care is part of Australia's Early Childhood Education and Care sector. For the purpose of this report, Child Care Management System (CCMS) data was used to identify whether a child had attended child care and their enrolled hours. Government-funded preschool attendance is not captured in the CCMS data.

A child is considered to have attended formal child care if they had a record in the CCMS prior to 2018 (their first year of school), meaning they were registered for subsidised child care and were charged for a session. As such, the results consider child care attendance at any age, not just in the year before school.

There are limitations with this definition: preschool attendance in fully state government-funded preschools is not captured; children with any attendance, even at a very low level, are captured; and families who attended a non-Child Care Benefit approved service, such as the former Budget Based Funded services, are not included (noting these groups are likely to be very small).

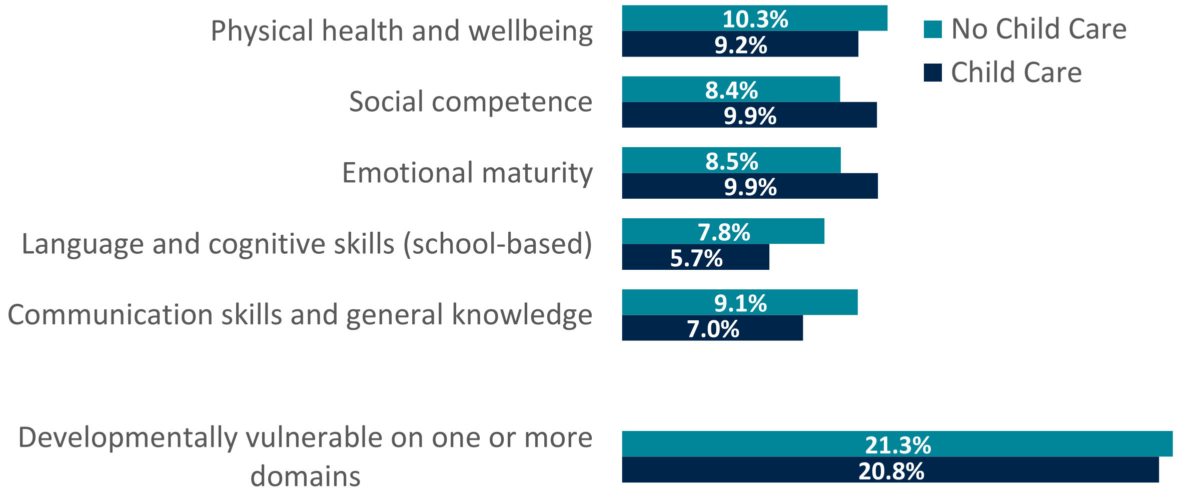
Charged hours are used as a proxy for attended hours in the analysis as there is no information about the actual hours of care attended by a child. Data custodians have suggested that, on average, charged hours represent approximately 70 per cent of attended hours. Therefore, it is likely that we will have overestimated attendance hours using this method.

Association with developmental vulnerabilities

Attending child care or preschool has a range of impacts on child development. The Effective Provision of Pre-School Education study found that a longer preschool (as defined in the United Kingdom) experience was associated with greater gains in cognitive ability (Sylva et al. 2004). Changing carers, be it through attending child care or a change in primary carer, can have short term effects on behaviour and distress, but these changes return to normal within weeks (Cryer et al. 2005). Some studies have found that the longer a child spends in child care, the more likely children are to fall ill or display behavioural issues (NICHD Early Child Care Research Network 2005; Belsky et al. 2007). However, Rey-Guerra et al. (2023) used a more robust within-child design – exploiting variations in the hours used by the same child, rather than comparing different children who attended for different amounts of hours – and found that increases in centre-based care hours were not associated with changes in child behaviour across five countries. Studies have also observed that learning outcomes of children (and in particular boys) who attended any child care before the age of three are better than those of children who attended no child care at all (Kalb et al. 2014).

Figure 1 shows the rates of developmental vulnerability on one or more domains (DV1), as well as rates of developmental vulnerability on each of the five domains by formal child care attendance. Children who attended child care had an overall lower rate of vulnerability than children who did not. Looking at the individual domains, children who attended formal child care had lower rates of vulnerability in the physical health and wellbeing, language and cognitive skills (school-based), and communication skills and general knowledge domains. However, they exhibited higher rates of developmental vulnerability in the social competence and emotional maturity domains. The following fact sheets explore this further by considering the quality of child care and the amount of hours used.

Figure 1. Proportion (%) of children who are developmentally vulnerable, by domain, by formal child care attendance, 2018 AEDC cohort.



Source: Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.

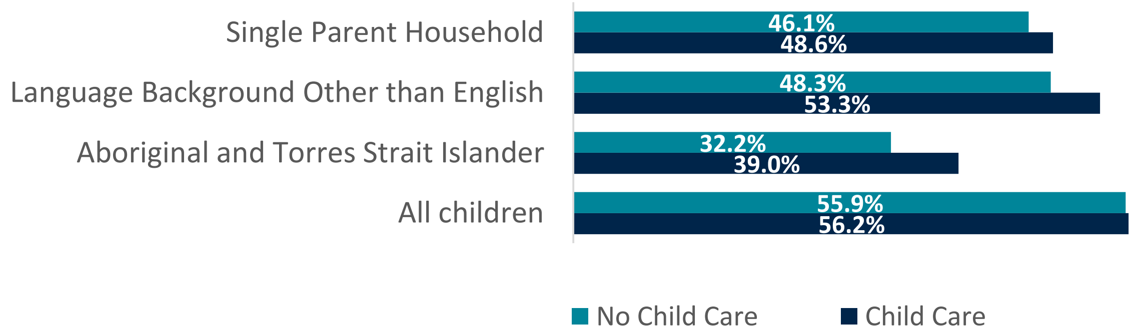
Notes: This figure compares children from the 2018 cohort of the Australian Early Development Census. DV1: Developmentally vulnerable on one or more domains. N = 272,626. Child care attendance is identified through the child having one or more records in the CCMS prior to 2018. Children in both the No Child Care or the Child Care groups may have separately attended preschool.

Association with being developmentally on track for selected equity groups

In Figure 2, the proportion of children who are developmentally on track on all 5 domains is shown for all children split by equity groups, and by their formal child care attendance. For the whole AEDC cohort, children who attended formal child care had a higher rate of being developmentally on track on all 5 domains, consistent with the lower DV1 rate for the child care attendance group shown in Figure 1.

Though the rate of being developmentally on track on all domains is lower for the equity groups compared to the rate for all children, attending formal child care is associated with higher rates of being developmentally on track on all domains for Aboriginal and Torres Strait Islander children, children from a single parent household and those with a language background other than English, consistent with other studies (AIHW 2015). For Aboriginal and Torres Strait Islander children who attended child care, there is a 14 percentage point increase in the rate of being developmentally on track on all 5 domains, while the increase is around 5 and 2.5 percentage points for children from other language backgrounds and single parent households respectively.

Figure 2. Proportion (%) of children who are developmentally on track on all 5 domains, by equity group, by formal child care attendance, 2018 AEDC cohort.



Source: Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.

Notes: This figure compares children from the 2018 cohort of the Australian Early Development Census. Child care attendance is identified through the child having one or more records in the CCMS quarterly enrolment data prior to 2018. Children in both the No Child Care or the Child Care groups may have separately attended preschool. On track on all 5 domains measured differently to currently used AEDC methods (see Methodology).

Child care attendance drivers

To add context to our findings we investigated the relationships between parental and child-centric factors and the rates of formal child care attendance. The analysis is descriptive and is therefore limited in its ability to indicate any causal links between these factors and child care attendance.

Child care attendance rates also differ by remoteness of the areas. Figure 3 shows that children living in metropolitan areas had the highest child care attendance rates, at 76 per cent. Children from very remote areas had the lowest rate at 48 per cent. This may be partly caused by limited availability of child care services in remote and very remote communities.

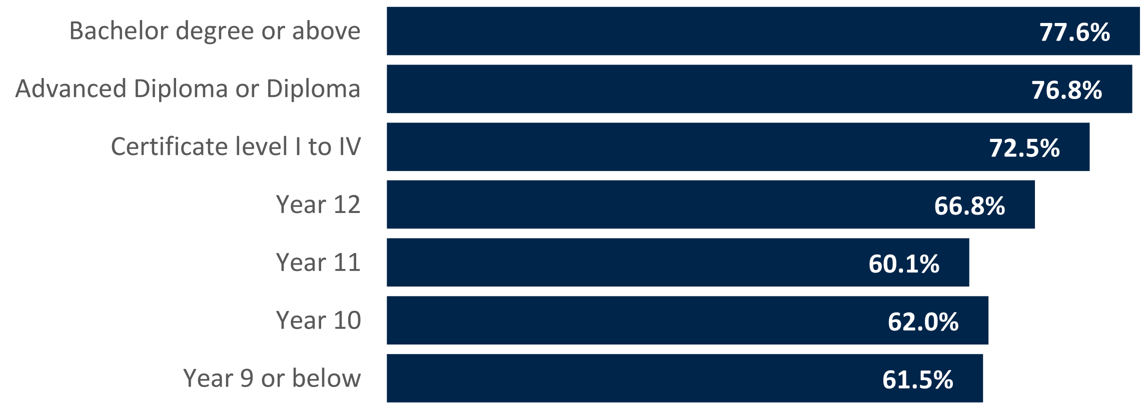
Figure 3. Proportion (%) of children who attend formal child care by remoteness, 2018 AEDC cohort.

Source: Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.

Notes: This figure compares children from the 2018 cohort of the Australian Early Development Census. N = 271,400. Only children with known remoteness status shown. Child care attendance is identified through the child having records in the CCMS prior to 2018. Remoteness identified through child's most remote address (see Methodology section).

Figure 4 shows the rate of formal child care attendance versus the highest educational attainment of the most educated parent or carer. Children with more educated parents were more likely to attend child care. This is consistent with parents with higher educational attainment having higher rates of employment and higher earning potential, both factors leading to higher demand for formal child care.

Figure 4. Proportion (%) of children who attend formal child care, by highest educational attainment of most educated parent, 2018 AEDC cohort.



Source: Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.

Notes: This figure compares children from the 2018 cohort of the Australian Early Development Census. N = 260,233. Only children with known parental highest educational attainment are shown. Child care attendance is identified through the child having records in the CCMS prior to 2018. Parent or carer education is identified using the AEDC (see *Methodology* section).

Figure 5 compares the rates of attendance for children in single parent households with children who are in dual parent households. There was only a small difference in child care attendance rates.

Figure 5. Proportion (%) of children who attend formal child care, by household status, 2018 AEDC cohort.

Source: Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.

Notes: This figure compares children from the 2018 cohort of the Australian Early Development Census. N = 267,580. Only children with known single-parent household status shown. Child care attendance is identified through the child having records in the CCMS prior to 2018. Single-parent household status identified through child and parent(s) shared address data (see *Methodology* section).

Preschool

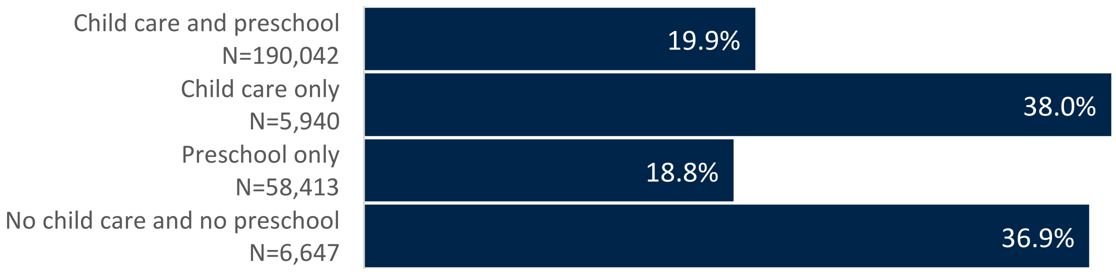
Preschool services deliver a structured play-based learning preschool program. According to formal requirements under the former Universal Access National Partnership, services were to be delivered by a qualified teacher (in practice, having 'access to' an early childhood teacher was accepted for services to meet the National Quality Standards). Preschool programs are aimed only at children in the year or two before they start full time school. Preschool starting ages differ by state or territory (Productivity Commission 2023). Children can attend dedicated preschool services attached to a school or standalone services, or through a preschool program at a child care centre.

Preschool attendance was identified either through the preschool attendance flag in the AEDC, or if a child had attended 600 hours of Long Day Care (LDC) child care based on the CCMS in their year before school. Attendance for 600 hours at LDC in the year before school corresponds to the Universal Access National Partnership aim from both federal and state and territory governments to guarantee 600 hours of preschool or high-quality child care to all children (DESE 2021). Measuring child care attendance using CCMS data does not allow us to distinguish between preschool programs in a child care centre and child care without a preschool program. We are also unable to capture attendance at fully state government-funded preschools using administrative data, although this should be captured in preschool attendance as reported in the AEDC.

We observed that children who attended preschool had far lower rates of developmental vulnerability on one or more domains compared to those who did not (Figure 6). Most children attended both child care and preschool, with a rate of developmental vulnerability of 19.9%. The preschool only group (identified as attended preschool from the AEDC but not in the CCMS data) has a similar rate of developmental vulnerability of 18.8%.

For the group who did not attend preschool (no matter whether they had attended formal child care) there is a significant risk of developmental vulnerability of close to one in three, compared to 1 in five for those who attended preschool. Around 4.8 per cent of the population do not attend preschool, with the apparent high rate of developmental vulnerability in this group potentially reflecting other disadvantages in life. Finding out why this small group of children do not attend preschool and helping them access preschool could reduce their rates of developmental vulnerability.

Figure 6. Proportion (%) of children who are developmentally vulnerable on one or more domains, by preschool and formal child care attendance, 2018 AEDC cohort.



Source: Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.

Notes: This figure compares children from the 2018 cohort of the Australian Early Development Census. N = 261,042. Only children with known preschool attendance status shown. Preschool attendance was identified through AEDC data or 600 hours of Long-Day Care attendance as per CCMS in year before school (see *Methodology* section).

References

AIHW (Australian Institute of Health and Welfare) (2015) *Literature review of the impact of early childhood education and care on learning and development*, catalogue number CWS 53, AIHW, Australian Government, accessed 3 September 2024.

Belsky J, Vandell DL, Burchinal M, Clarke-Stewart KA, McCartney K, Owen MT, and NICHD Early Child Care Research Network (2007) 'Are there long-term effects of early child care?', *Child development*, *78*(2):681–701, doi: 10.1111/j.1467-8624.2007.01021.x.

Cryer D, Wagner-Moore L, Burchinal M, Yazejian N, Hurwitz S and Wolery M (2005) 'Effects of Transitions to New Child Care Classes on Infant/Toddler Distress and Behavior', *Early Childhood Research Quarterly*, 20(1):37–56, doi:10.1016/j.ecresq.2005.01.005.

DESE (Department of Education, Skills and Employment) (2021) *Universal Access National Partnership*, DESE website, accessed December 22 2021.

Kalb G, Tabasso D and Zakirova R (2014) *'Children's participation in early childhood education and care, and their developmental outcomes by Year 5: A comparison between disadvantaged and advantaged children'* Report for the Department of Education, Employment and Workplace Relations, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

NICHD (National Institute of Child Health and Human Development) Early Child Care Research Network (2005) *Child care and child development: Results from the NICHD study of early child care and youth development*, Guilford Press.

Productivity Commission (2023) Report on Government Services 20203 – Section 3 Early childhood education and care, Productivity Commission, accessed January 2024.

Rey-Guerra, C, Zachrisson, HD, Dearing, E, Berry D, Kuger S., Burchinal MR., Nærde, A, van Huizen T, and Côté, SM (2023) 'Do more hours in center-based care cause more externalizing problems? A cross-national replication study', *Child development*, 94(2):458–477, doi: 10.1111/cdev.13871

Sylva K, Melhuish E, Sammons P, Siraj-Blatchford I and Taggart B (2004) *The Effective Provision of Pre-school Education (EPPE) Project: final report,* Institute of Education, University of London, doi:10.1177/1476718X10387900.

The First Five Years: What makes a difference?

3.2 Quality of child care

Key findings

In this section we describe the associations between child care quality and developmental outcomes for different subgroups of the population.

- Children who attended above standard quality child care had lower rates of developmental vulnerability than children attending services assessed as not yet at standard quality child care.
- Children who attended preschool were more likely to be developmentally on track on all domains than those who did not attend preschool, and this held for children attending any given quality of child care.
- For all Socio-Economic Indexes for Areas (SEIFA) quintiles, children who attended above standard quality child care were more likely to be developmentally on track on all domains than those who attended not yet at standard quality child care.
- The difference in rates of being on track on all domains between SEIFA quintiles was generally larger than the difference in rates of being on track on all domains between child care quality categories within a SEIFA quintile. Children from the most advantaged SEIFA quintile who attended not yet at standard quality child care had lower rates of developmental vulnerability than those from least advantaged SEIFA quintile attending above standard quality child care.
- When the analysis compared all children from the same SEIFA quintile, children who attended preschool but not formal child care generally had the highest rate of being developmentally on track on all domains. For children from equity groups including Aboriginal and Torres Strait Islander children, children from a language background other than English (LBOTE) and children from single parent households, those who attended above standard quality child care had the highest rate of being developmentally on track on all domains.

Access to high quality child care has been shown to improve developmental outcomes in children. Carers with higher qualifications and lower child-carer ratios have been linked to improved cognition, language, and social skills in the first three years (Vandell and Wolfe 2000, Yamauchi and Leigh 2011).

The National Quality Standard (NQS) sets a high benchmark for early childhood education and care and outside school hours care services in Australia. Services are assessed and rated by their regulatory authority against the NQS and given a rating for each of the 7 quality areas and an overall rating based on these results (ACECQA 2021).

We created a quality factor to simplify the overall quality rating into four categories (see *Methodology* section for details):

- not yet at standard: overall rating of Significant Improvement Required or Working Towards National Quality Standard (NQS)
- at standard: overall rating that met the NQS
- above standard: overall rating of Excellent or Exceeding NQS
- provisional: child care quality not assessed by the NQF.

For each child, we identified child care quality as the most frequently occurring, non-missing, quarterly quality rating of all child care service providers attended by a child.

Figure 1 shows the rates of developmental vulnerability on one or more domains versus the quality of child care. We found that children who attended at standard services showed the same rates of developmental vulnerability as children who attended no child care. Children who attended not yet at standard child care had the highest rates of developmental vulnerability and children who attended above standard child care had the lowest rates. Quality indicators were not available for all child care providers; approximately 20 per cent of children attended child care with a provisional rating (see *Methodology* section for more details).

Figure 1. Proportion (%) of children who are developmentally vulnerable on one or more domains, by child care quality, 2018 AEDC cohort.

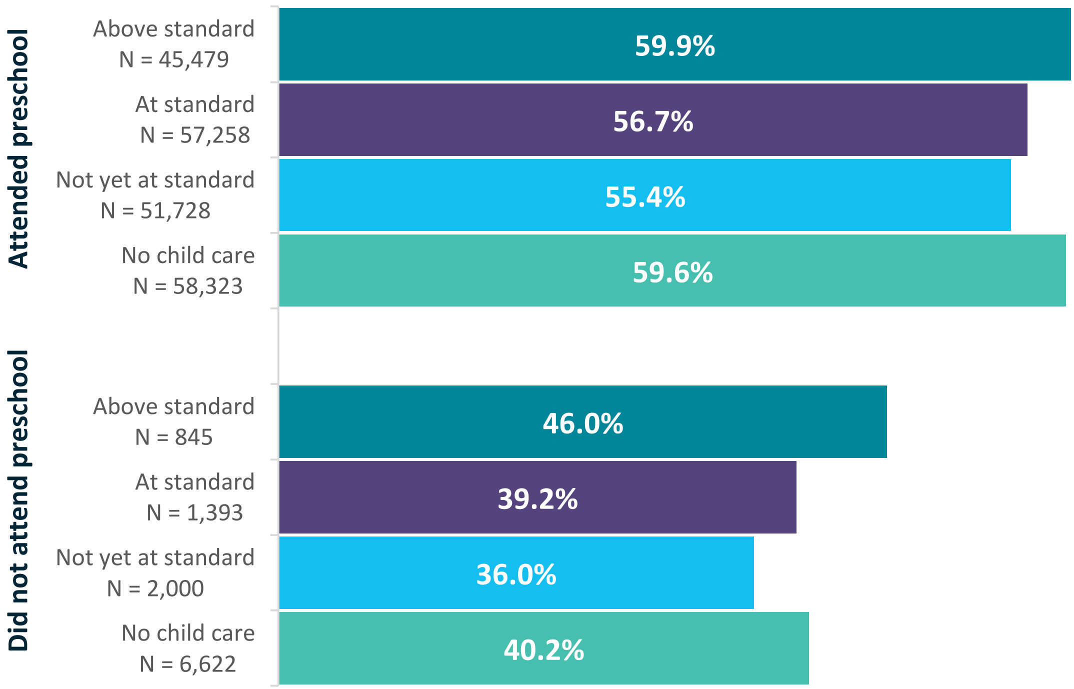
Source: Customised ‘First Five Years’ extract from the Multi-Agency Data Integration Project.

Notes: This figure compares children from the 2018 cohort of the Australian Early Development Census. N = 272,626. Children in all of the groups may have separately attended preschool..

Preschool and quality of child care

Figure 2 shows the rates of children developmentally on track on all domains versus the overall quality of the child care in a child's history of child care attendance, split by preschool attendance. As reported in chapter 3.1, a small proportion of children (4.8%) did not attend preschool.

Figure 2: Proportion (%) of children who were on track on all domains, by overall child care quality and preschool attendance, 2018 AEDC cohort.



Source: 'First Five Years' extract from the Multi-Agency Data Integration Project.

Notes: This figure compares children from the 2018 cohort of the Australian Early Development Census. The size of the cohort is slightly different to that on DV1 due to the differing requirements of having a valid "DV1" score or "on track on all domains" score. Only children with known preschool attendance status are shown. Preschool attendance was identified through AEDC data or 600 hours of Long-Day Care attendance as per CCMS in year before school (see *Methodology* section).

Children who attended preschool and either above standard quality child care or no child care were the most likely to be developmentally on track on all domains.

Children who attended higher quality child care were more likely to be developmentally on track on all domains, regardless of whether they attended preschool or not. However, the effect of different child care quality varies according to preschool attendance. For children who attended preschool the difference in rates of being developmentally on track on all domains between above standard and not yet at standard quality child care was approximately 5 percentage points, whereas for children who did not attend preschool the difference was approximately 10 percentage points (Figure 2).

Children who attended preschool were more likely to be developmentally on track on all domains for all quality categories than children in the same quality category who did not attend preschool. Children who attended child care but not preschool were most likely to attend not yet at standard quality child care (with the not yet at standard group having the largest attendance of around 2,000 children). This subgroup of children was the least likely to be developmentally on track on all domains.

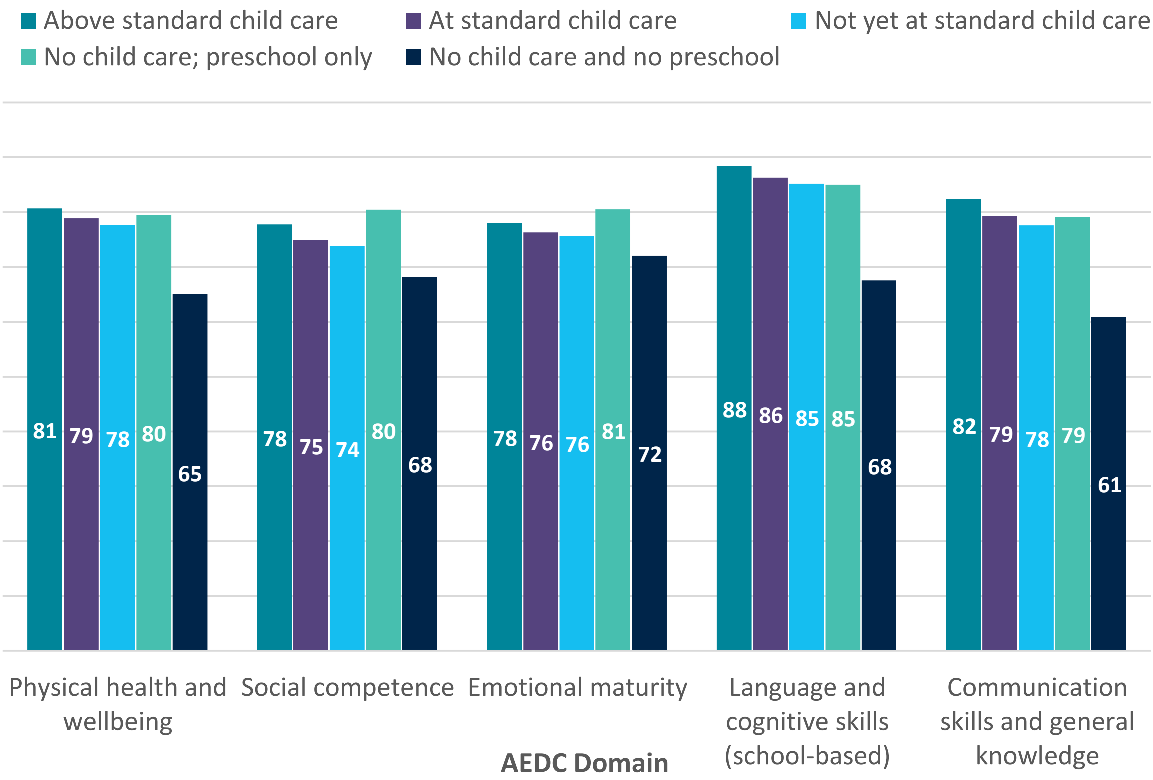
Association between child care quality and developmental domains

For each AEDC domain, children who attended above standard quality child care were more likely to be developmentally on track than those who attended not yet at standard quality child care (Figure 3). The largest difference between rates for above standard and not yet at standard quality ratings occurred for the communication skills and general knowledge, and social competence domains (4 percentage points), while the lowest difference occurred for the emotional maturity domain (2 percentage points).

For all AEDC domains, children who attended neither child care nor preschool were the least likely to be developmentally on track. The largest differences were observed in the cognitive domains (language and cognitive skills (school-based) and communication skills and general knowledge), and in physical health and wellbeing.

Children who attended above standard child care had the highest rate of being developmentally on track on all domains (Figure 2). When considering each developmental domain separately, the results change between domains. Children who attended preschool but not formal child care were the most likely to be developmentally on track in the social competence and emotional maturity domains. For the cognitive domains (language and cognitive skills (school-based) and communication skills and general knowledge), children attending above and at standard child care (and who may or may not have attended preschool) had the highest rate of being developmentally on track.

Figure 3: Proportion (%) of children who were developmentally on track, by AEDC domain and overall child care quality, 2018 AEDC cohort.



Source: 'First Five Years' extract from the Multi-Agency Data Integration Project.

Notes: This figure compares children from the 2018 cohort of the Australian Early Development Census. Only children with known preschool attendance status are shown. Preschool attendance was identified through AEDC data or 600 hours of Long-Day Care attendance as per CCMS in year before school (see *Methodology* section).

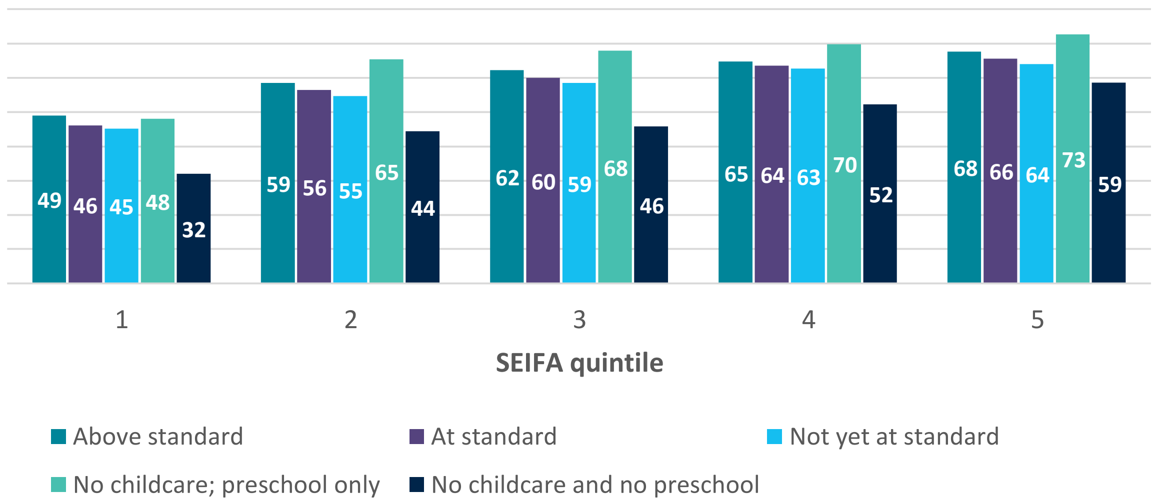
Child care quality and area-based socio-economic status

To describe the socio-economic status of a child's area of residence, we used one of the ABS' SEIFA measures – the Index of Relative Socio-Economic Advantage and Disadvantage (IRSAD) (ABS 2018). SEIFA quintiles were used, where quintile 5 represents the areas with the highest socio-economic status and quintile 1 represents the lowest. By splitting the cohort in this way, the analysis is effectively controlling for differences between SEIFA quintiles.

For all SEIFA quintiles, children who attended above standard quality child care were more likely to be developmentally on track on all domains than those who attended not yet at standard quality child care. The difference in rates between high and low quality child care was similar across SEIFA quintiles (2-4 percentage points). For all SEIFA quintiles, children who attended neither preschool nor child care were the least likely to be developmentally on track on all domains.

For all quality categories, rates of being developmentally on track on all domains increased as the SEIFA quintile increased. Even the lowest rate for SEIFA quintile 5 (59% for no child care and no preschool) was greater than the highest rate for SEIFA quintile 1 (49% for above standard child care).

Figure 4: Proportion (%) of children who were developmentally on track on all domains, by household SEIFA quintile and overall child care quality, 2018 AEDC cohort.

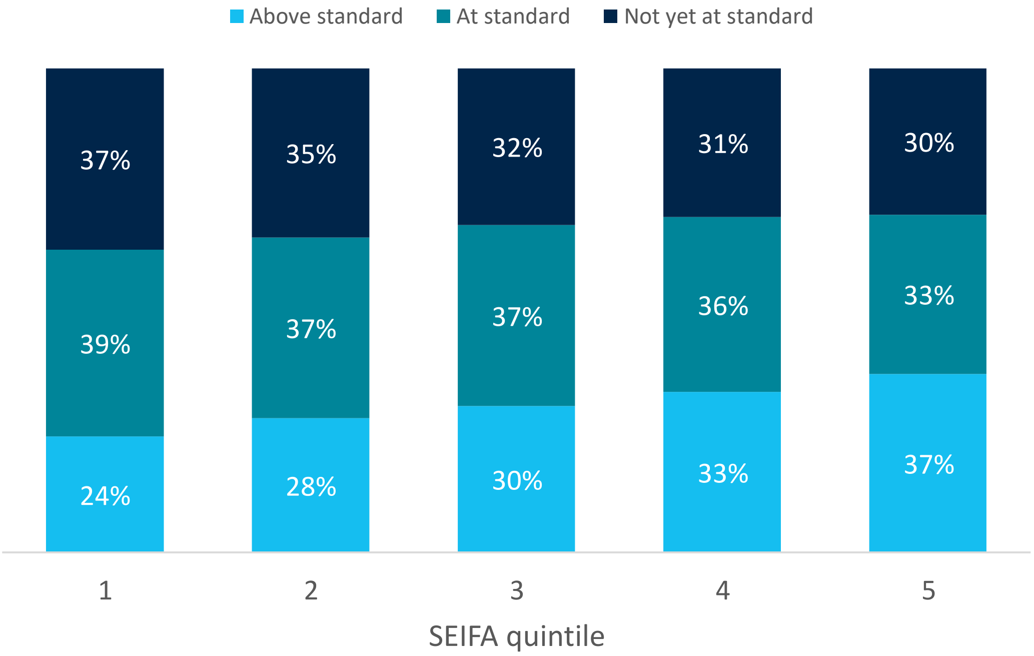


Source: 'First Five Years' extract from the Multi-Agency Data Integration Project.

Notes: This figure compares children from the 2018 cohort of the Australian Early Development Census. Only children with known preschool attendance status and SEIFA quintile are shown. Preschool attendance was identified through AEDC data or 600 hours of Long-Day Care attendance as per CCMS in year before school (see *Methodology* section).

Children from lower SEIFA quintiles were more likely to attend not yet at standard quality child care and were less likely to attend above standard quality child care (Figure 5). Only 24 per cent of children in SEIFA quintile 1 attended above standard quality child care, compared to 37 per cent of children in SEIFA quintile 5. Conversely, 37 per cent of children in the lowest SEIFA quintile attended not yet at standard child care, while only 30 per cent of children in the highest SEIFA quintile attended not yet at standard child care.

Figure 5: Proportion (%) of children who attended above, at, and not yet at standard quality child care, by household SEIFA quintile, 2018 AEDC cohort.



Source: 'First Five Years' extract from the Multi-Agency Data Integration Project.

Notes: This figure compares children from the 2018 cohort of the Australian Early Development Census. Only children with known SEIFA quintile and known child care quality are shown.

Association between child care quality and developmental outcomes for equity groups

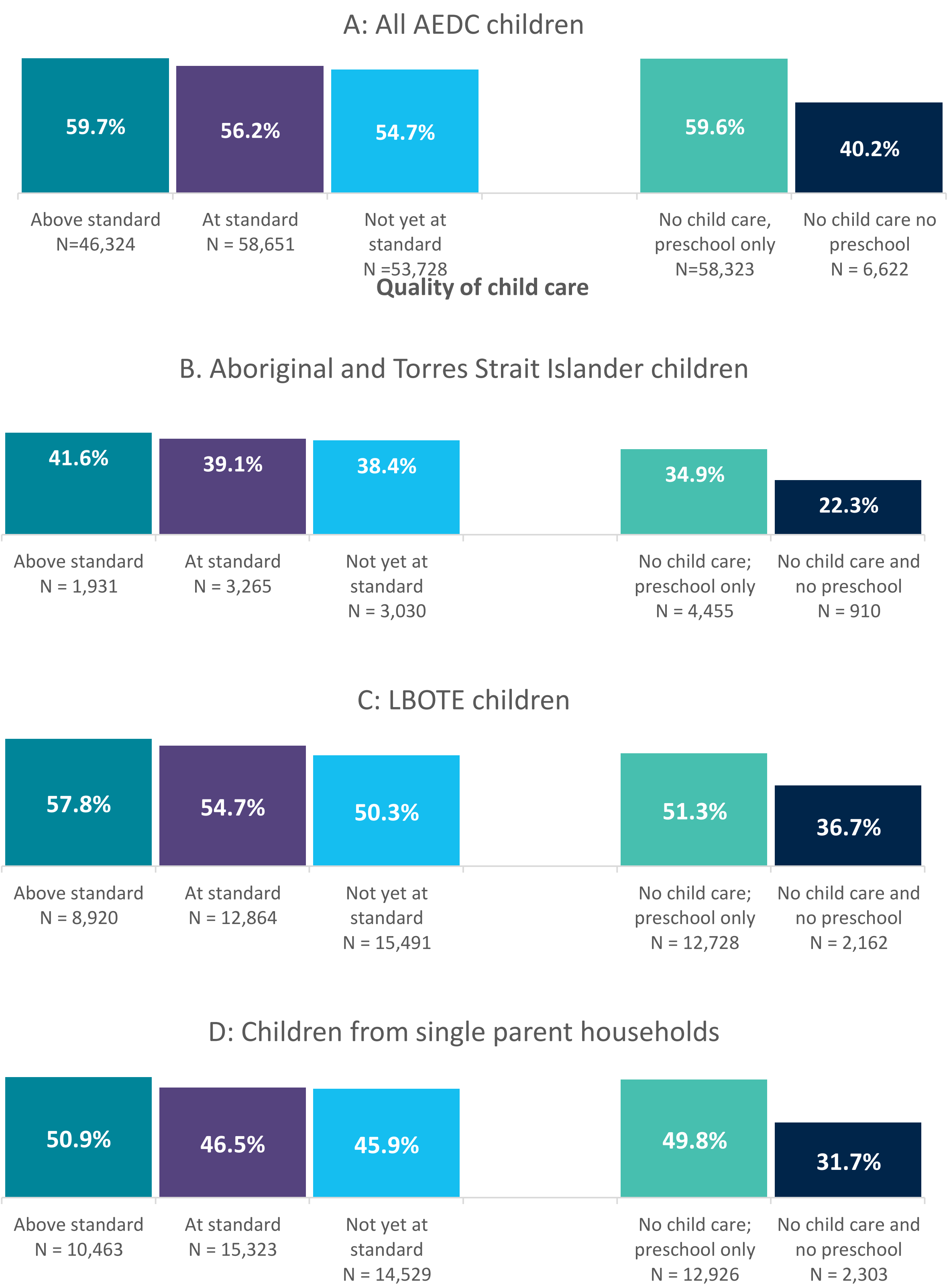
In this section, we looked at whether child care quality and preschool attendance have different associations for different cohorts: Aboriginal and Torres Strait Islander children, children from a language background other than English, and children from single parent households. The corresponding data for the entire AEDC cohort (Figure 2) is replotted here as Figure 6A for comparison.

For all groups, children attending higher quality child care were more likely to be developmentally on track on all domains (Figure 6). Rates of being on track on all domains were approximately 3 percentage points higher for Aboriginal and Torres Strait Islander children who attended above standard quality child care than for those whose child care was not yet at standard quality (Figure 6B). Children from a language background other than English who attended above standard child care had rates of being developmentally on track on all domains approximately 7.5 percentage points higher than those attending not yet at standard quality child care (Figure 6C). The difference is approximately 5 percentage points for children from single parent households who attended above standard quality child care compared to those who attended at or not yet at standard quality child care (Figure 6D).

For all equity cohorts, children who attended neither child care nor preschool were the least likely to be developmentally on track on all domains. Rates of being developmentally on track on all domains were nearly 20 percentage points higher for those who attended above standard quality child care, compared to those who attended neither child care nor preschool: for Aboriginal and Torres Strait Islander children the difference was 41.6 compared to 22.3 per cent, for children from a language background other than English: 57.8 compared to 36.7 per cent, and for children from single parent households: 50.9 compared to 31.7 per cent.

The situation is different when comparing the quality of the child care group with the preschool only group. Looking at all the AEDC children, the non-child care, preschool only group has only a slightly lower rate of being developmentally on track on all domains compared to those attending above standard child care (Figure 6A). When looking at the equity cohorts, the preschool only group has only a slightly lower rate of being developmentally on track on all domains compared to those attending above quality child care for the children from single parent households. For Aboriginal and Torres Strait Islander children, the preschool only group has a lower rate of being developmentally on track on all domains than those attending child care of any quality. Similarly for the cohort with a language background other than English, the preschool only group has a lower rate of being developmentally on track on all domains, only slightly above that of the children attending not yet at standard child care.

Figure 6: Proportion (%) of children who were developmentally on track on all domains, by quality of child care services attended, for Aboriginal and Torres Strait Islander children, children from Language background other than English, and children from single parent households, 2018 AEDC cohort.



Source: 'First Five Years' extract from the Multi-Agency Data Integration Project.

References

ABS (Australian Bureau of Statistics) (2018) *2033.0.55.001 – Census of Population and Housing: Socio-Economic Indexes for Areas (SEIFA), Australia, 2016*, ABS, accessed 5 January 2021.

ACECQA (Australian Children's Education and Care Quality Authority) (2021) *National Quality Standard*, ACECQA, accessed 10 December2021.

Vandell D and Wolfe B (2000) 'Child Care Quality: Does It Matter and Does It Need to Be Improved?', *Institute for Research on Poverty Special Report*, 78.

Yamauchi C and Leigh A (2011) 'Which children benefit from non-parental care?', *Economics of Education Review*, Elsevier 30(6):1468-1490.

The First Five Years: What makes a difference?

3.3 Hours of child care attendance

Key findings

In this section we describe the associations between child care hours and rates of being developmentally on track for different subgroups of the population.

- Children who were charged for an average of 5 to 30 hours of child care per week had lower rates of developmental vulnerability, and higher rates of being developmentally on track on all domains, than children with less or more hours of child care.
- Children with 3,000 to 7,000 total hours of child care before starting school had lower rates of developmental vulnerability than those with less or more total hours.
- Aboriginal and Torres Strait Islander and LBOTE children with 15 to less than 30 hours were the most likely to be developmentally on track on all domains, and those with over 30 hours had higher rates of being developmentally on track than those with less than 15 hours (including those with no formal child care hours).
- The lowest SEIFA and single parent cohorts had a similar pattern to the entire AEDC cohort, with children who attended preschool without child care having the highest rates of being developmentally on track.

Hours of child care usage

The analysis uses the average *charged* hours of child care per week captured in the Child Care Management System (CCMS), during the time that a child attended child care. The average weekly hours do not capture a variation in the patterns of child care use such as the age the child started in care, or how consistently a child attended child care across different years. This variation might be expected to influence the risk of developmental vulnerabilities. To account for the consistency in the use of child care, we look at the total hours a child was charged across the years before attending school. Total child care hours were calculated as the sum of all charged hours of child care prior to 2018 where the child was registered in the CCMS. Combining the average weekly hours and the total hours provides some visibility of the complex interplay between the total time spent in child care and the intensity of child care used.

Actual hours of child care attended was not available for the 2018 AEDC cohort, noting that hours attended are generally fewer than hours charged (see Figure 1 in the *Appendix* for a comparison between hours attended and charged for the 2021 AEDC cohort). Children not registered in the CCMS were classified as having zero average weekly child care hours.

For simplicity, in some of the analysis we categorised children into three groups based on their hours of child care – less than 15, 15 to less than 30, and 30 or more average charged hours per week.

Children who never attended formal child care were further categorised into two groups – attending and not attending preschool. Children were considered to have attended preschool but not child care if they were flagged as attending preschool in the AEDC but were not registered in the CCMS. Children were considered to have attended neither preschool nor child care if they were flagged as not attending preschool in the AEDC and were not registered in the CCMS.

Note government-funded preschool attendance is not captured in the CCMS data, so early childhood education and care service attendance may be understated particularly for those with lower average charged hours.

For details of the average weekly charged hours, see *Methodology* section.

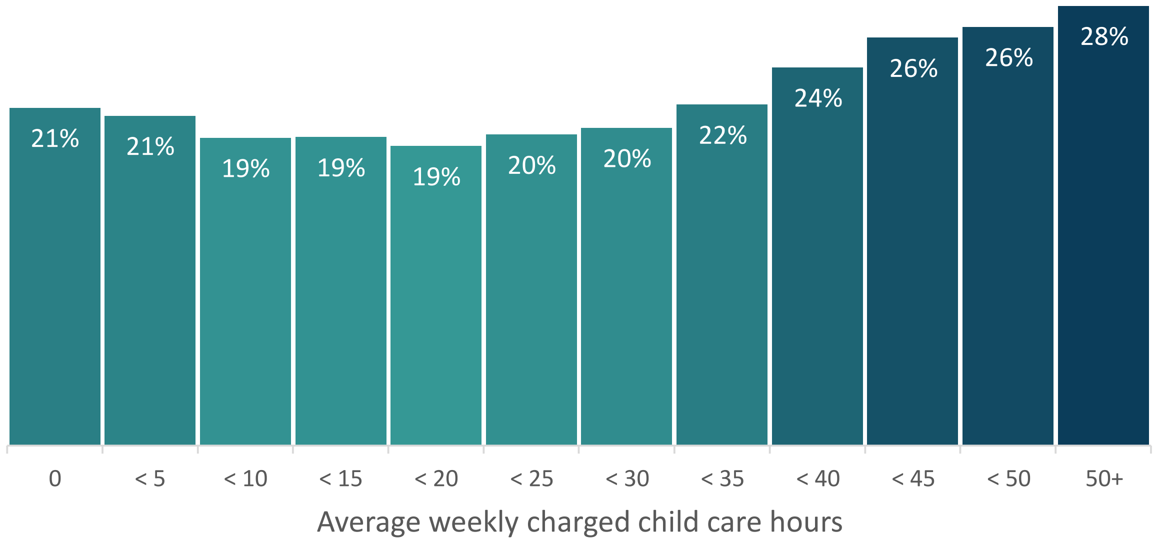
Hours of child care and developmental vulnerability

Hours of child care attendance have been found to be connected to children’s development. Some studies have found that children who attend child care for longer hours display more behavioural problems and minor illnesses (NICHD 2005; Belsky et al. 2007). Studies have also observed that learning outcomes of children (and in particular boys) who attended any child care before the age of three are better than those of children who attended no child care at all (Kalb et al. 2014).

This section presents some results based on the summary measure of developmental vulnerability in one or more AEDC domains (DV1). These should be considered along with developmental outcomes on the individual domains shown later, as results for individual domains can diverge from those for summary measures.

Figure 1 shows that rates of developmental vulnerability in one or more AEDC domains (DV1) initially decline and then increase as average weekly child care hours increase. Children with 5 to 25 average weekly charged hours of child care had the lowest rates of developmental vulnerability.

Figure 1. Proportion (%) of children that are developmentally vulnerable on one or more domains (DV1), by average weekly charged child care hours, 2018 AEDC cohort.



Source: Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.

Notes: This figure compares children from the 2018 cohort of the Australian Early Development Census. N = 272,626. Each child in the dataset contributes to only the minimum bin that describes their weekly hours. Children may have separately attended preschool, which was not captured in the CCMS data.

As children were charged for more than 30 hours in child care their rates of developmental vulnerability increased. Children who were charged for between 30 and 40 hours per week had similar rates of developmental vulnerability to children who attended fewer than five hours per week. Children who were enrolled for more than 40 hours per week had the highest rates of developmental vulnerabilities.

Note that results in this paper focus only on child care hours, rather than the attendance of Early Childhood Education and Care (ECEC) in general, since preschool attendance data is not available. In section 3.1 Figure 6, it is shown that children who attend preschool have a lower rate of being developmentally vulnerable than those who do not.

Comparing outcomes between different groups of children who are using different hours of child care is not straightforward, as any difference in outcomes can be due to other differences between the groups rather than just their child care use. This is the case even if comparisons adjust for variations in some other observed characteristics between the groups. The Productivity Commission (PC 2024) caution that it is harder to find evidence of risks related to high hours of ECEC in research that can more credibly isolate the effects of this intensity. The analysis given here adds to the evidence considered in the PC report, noting the results include non-cognitive as well as cognitive domains; focus on child care rather than preschool programs aimed at older children; and are for an entire cohort rather than being based on specialised programs targeted at disadvantaged children.

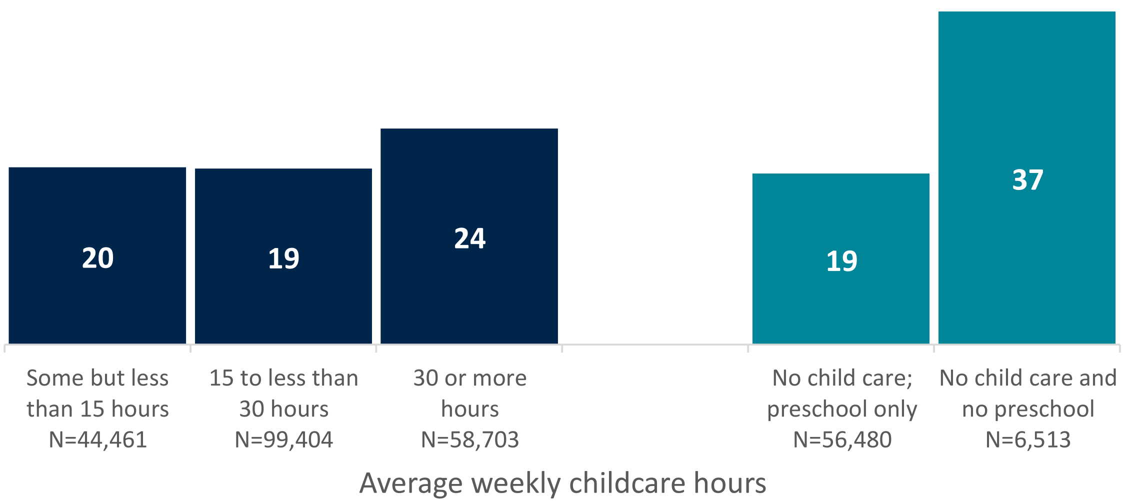
Pattern of child care attendance

This section focusses on the effect of child care hours on outcomes. Other analysis that considered the combined effects of quality and hours suggests that these dimensions can be considered separately, rather than there being significant interaction effects.

To summarise weekly formal child care, we categorised children attending child care into three groups based on their hours of child care – less than 15, 15 to less than 30, and 30 or more average charged hours per week, where the average was taken over the quarters a child was using formal child care. Children who never attended formal child care were further split into two groups, depending on whether or not they had attended preschool.

Children using formal child care with less than 30 average weekly charged child care hours (Figure 2) had very similar DV1 rates (approximately 20 per cent); this increased to 24 per cent for children with longer child care hours. Children who attended preschool but not child care had similar DV1 rates to those with less than 30 hours (19 per cent), while those who attended neither had a substantially higher rate (37 per cent). The small group of children who attended neither preschool nor child care were almost twice as likely to be developmentally vulnerable on one or more domains as other children with no or less than 30 hours of child care.

Figure 2: Proportion (%) of children who were developmentally vulnerable on one or more domains (DV1), by average weekly child care hours, 2018 AEDC cohort.



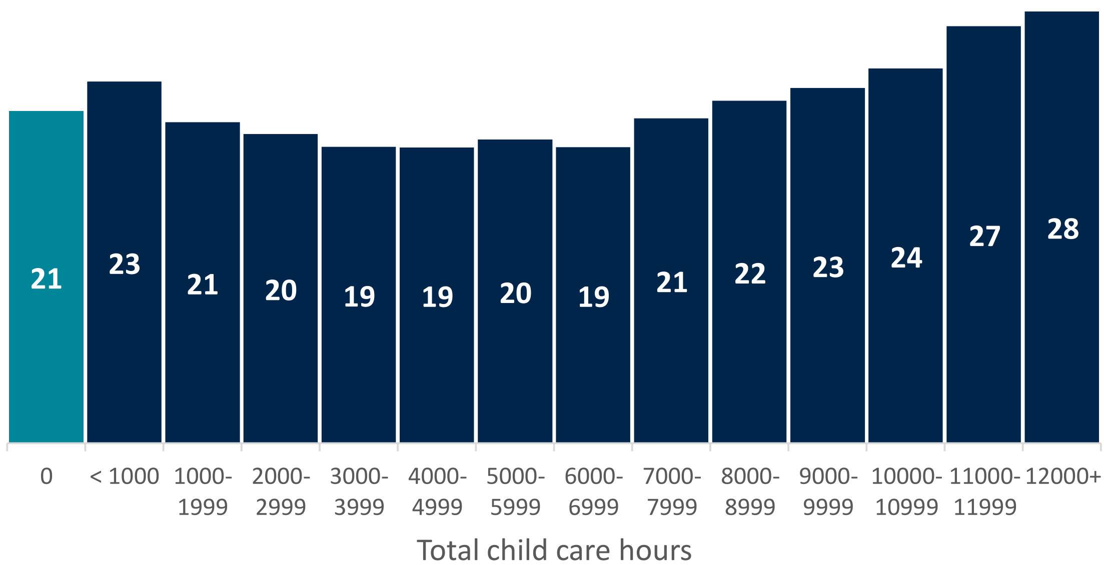
Source: Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.

Total hours in child care

As well as average hours charged per week, we also looked at the total charged hours across the years before attending school.

Children with 3,000 to 7,000 total child care hours had the lowest rates of developmental vulnerability (DV1, approximately 19 per cent). Those with some but fewer than 1,000 hours of child care had a higher rate of developmental vulnerability on one or more domains (23 per cent), noting that this could reflect underlying factors that contribute to less child care use and higher rates of developmental vulnerability (such as ill health). As the total number of hours increased beyond 7,000, rates of developmental vulnerability on one or more domains began to increase to a maximum rate of 28 per cent for children with 12,000 or more total child care hours.

Figure 3: Proportion (%) of children who were developmentally vulnerable on one or more domains (DV1), by total child care hours, 2018 AEDC cohort.



Source: Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.
Note: Children may have separately attended preschool, which was not captured in the CCMS data.

The heat map in Figure 4 (a) show rates of developmental vulnerability on one or more domains (DV1) for combinations of average weekly child care hours and total child care hours until the child attended school. The lowest rates of developmental vulnerability on one or more domains were found for those with the middle range of total hours (2,000-8,000 hours), and a middle range of average weekly hours (10-29 hours). The majority of children in our cohort were using child care in these ranges; see Figure 4(b) for the number of children in each bin. Further research could look at how outcomes vary depending on the age of the child when child care use occurred, separately to the number of hours.

Figure 4: (a) Heatmap of proportion (%) of children who were developmentally vulnerable on one or more domains (DV1), by total child care hours and average weekly hours, 2018 AEDC cohort.

Total hours	Average weekly hours <10	10-19	20-29	30-39	40-49	50+	Total
1-2,000	21%	22%	27%	32%	40%		19%
2,001-4,000		17%	22%	28%	26%	35%	18%
4,001-6,000		13%	18%	25%	29%	26%	18%
6,001-8,000			16%	21%	28%		19%
8,001-10,000				21%	25%	27%	21%
10,001-12,000				19%	25%	30%	24%
12,000+					29%	28%	26%
Total	21%	19%	20%	23%	26%	29%	

Notes: Cells with fewer than 100 children have been removed.

(b) Heatmap of number of children, by total child care hours and average weekly hours, 2018 AEDC cohort.

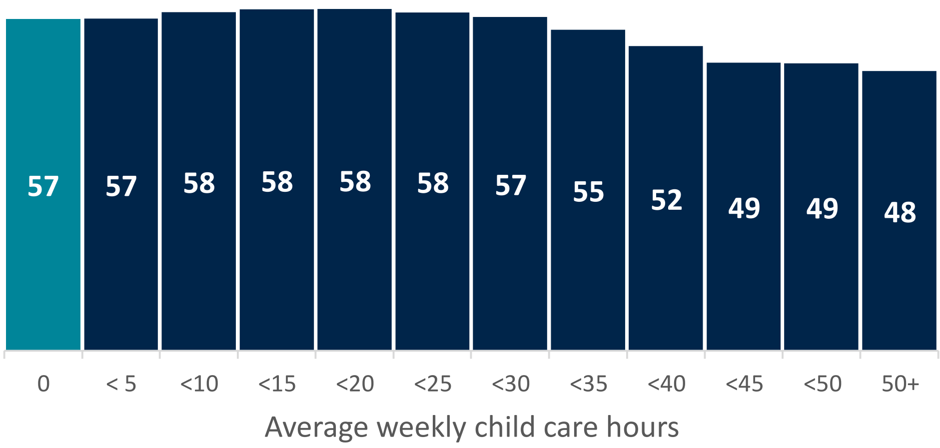
Total hours	Average weekly hours <10	10-19	20-29	30-39	40-49	50+
1-2,000	13456	25517	5933	877	101	
2,001-4,000	56	19839	21416	3124	571	1569
4,001-6,000		2012	25841	7631	1364	2137
6,001-8,000			6447	15736	2845	80
8,001-10,000			56	7117	6284	168
10,001-12,000				397	4780	269
12,000+					846	563

Source Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.

Child care hours and rates of being developmentally on track on all domains

Results in previous sections considered the rate of being developmentally vulnerable on one or more domains (DV1). Another summary indicator, the rate of being developmentally on track on all domains, reflects children's developmental strengths and can identify where things are working well to support children's development.

Figure 5: Proportion (%) of children who were developmentally on track on all domains, by average hours of child care per week, 2018 AEDC cohort.



Source: Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.

Note: Children may have separately attended preschool, which was not captured in the CCMS data.

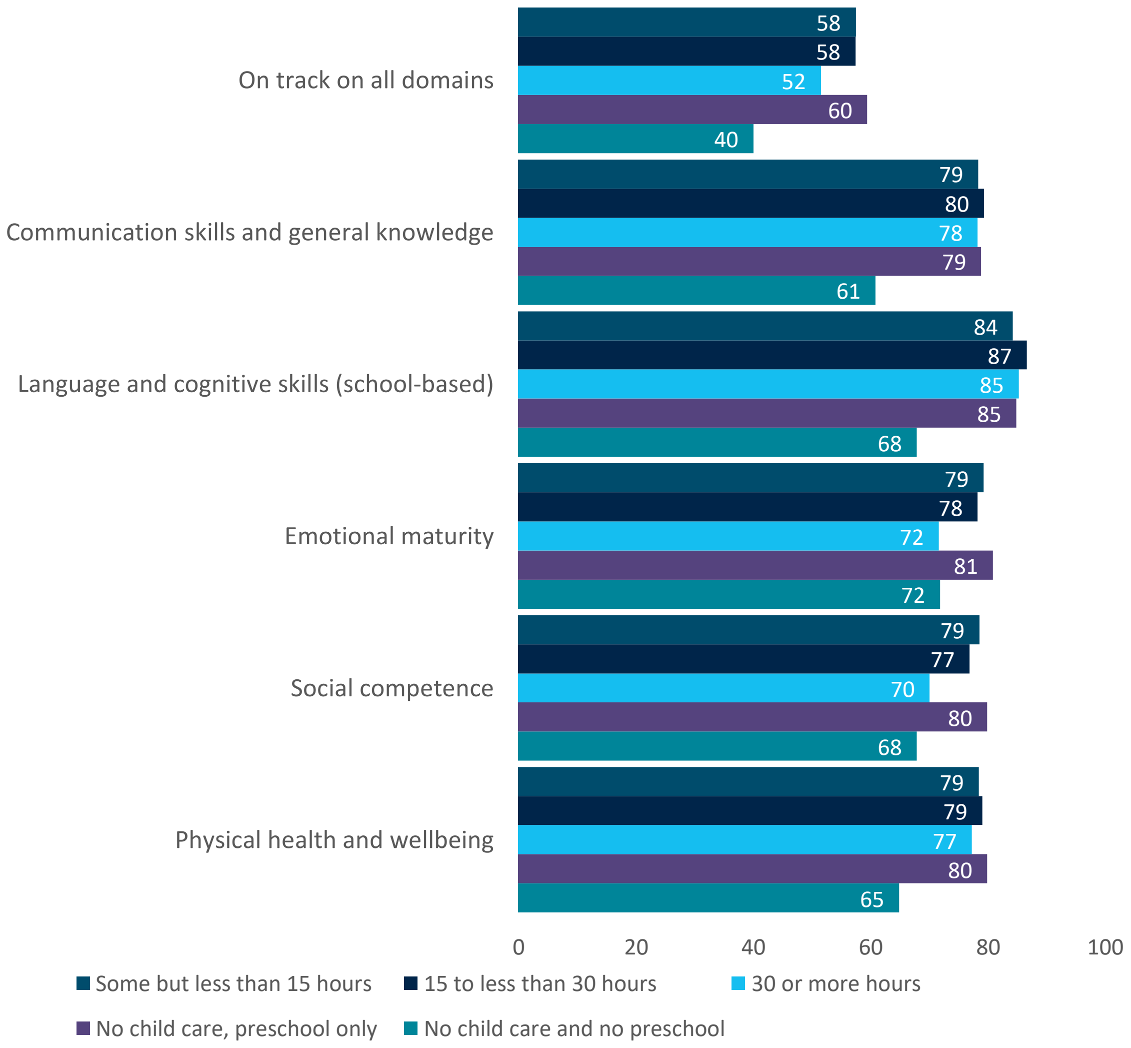
Children with fewer than 30 hours of average charged hours per week were the most likely to be developmentally on track on all domains (Figure 5). Rates of being developmentally on track on all domains declined for more than 30 hours, with a difference of 10 percentage points between the highest (58 per cent for 10-25 hours per week) and lowest (48 per cent for 50 or more hours per week) rates.

Hours of child care and being developmentally on track in each domain

In this section we look at rates of being developmentally on track on all domains and on each domain separately.

Children with some but fewer than 30 average weekly charged hours had the same rates of being developmentally on track on all domains (58 per cent in Figure 6), but children who had longer child care hours were less likely to be developmentally on track on all domains (52%). Children who attended preschool but not child care had high rates of being developmentally on track on all domains (60 per cent), while those who attended neither had the lowest rates (40 per cent).

Figure 6: Proportion (%) of children who were developmentally on track, by average weekly child care hours, for each domain 2018 AEDC cohort.



Source: Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.

Across the individual AEDC domains, the rate of being developmentally on track also differs with average child care hours.

Children who attended neither child care nor preschool had the lowest rate of being developmentally on track on all domains.

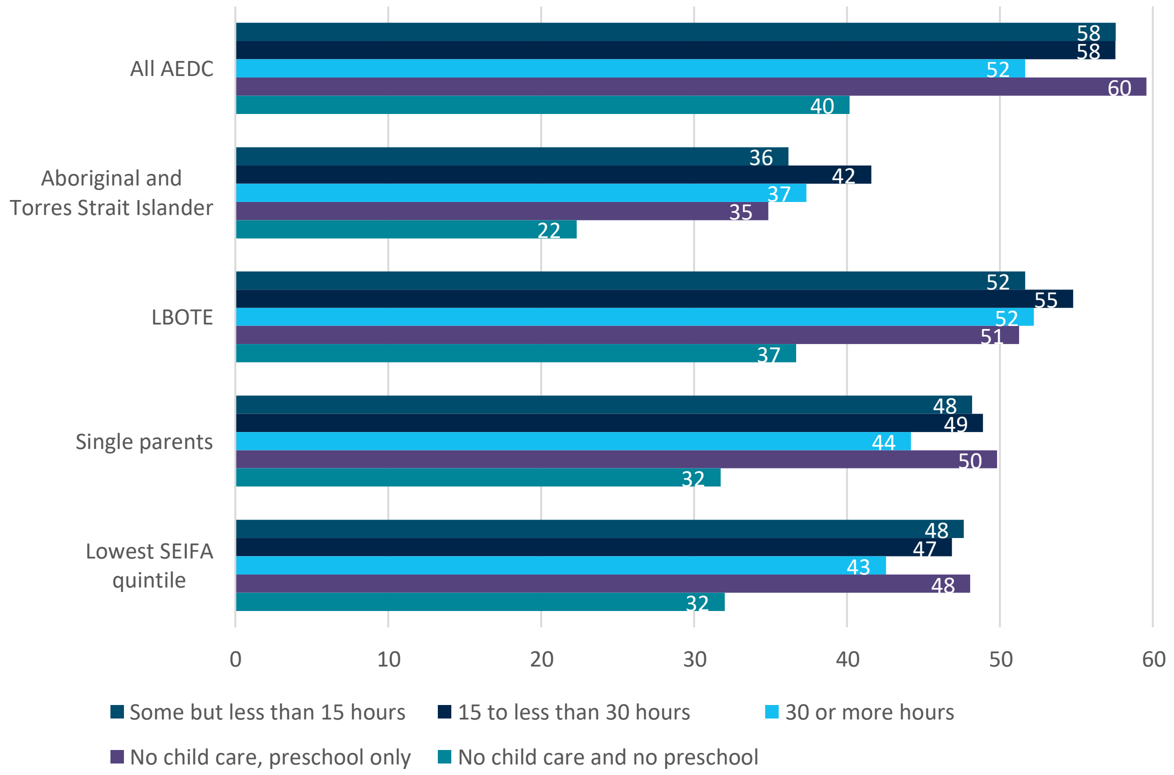
For the social competence and emotional maturity domains, children with 30 or more average charged hours had lower rates of being developmentally on track than those with fewer hours (with the difference reaching 9 and 7 percentage points for the social competence and emotional maturity domains, respectively).

Children with 15 to 30 average weekly charged hours were the most likely to be developmentally on track on the communication skills and general knowledge and language and cognitive skills (school-based) domains. However, the difference between the highest and lowest rates was smaller (around 1 to 3 percentage points). The physical health and wellbeing domain followed a broadly similar pattern.

Child care hours and development outcomes for equity groups

In this section, we look at average charged child care hours and their association with development outcomes for several equity groups including Aboriginal and Torres Strait Islander children, children from a language background other than English, children from single parent households, and children from the lowest SEIFA quintile. The rate of being developmentally on track on all domains for all AEDC children is also shown for comparison.

Figure 7: Proportion (%) of children who were developmentally on track on all domains, by average weekly child care hours, for all AEDC cohort and different equity groups.



Source Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.

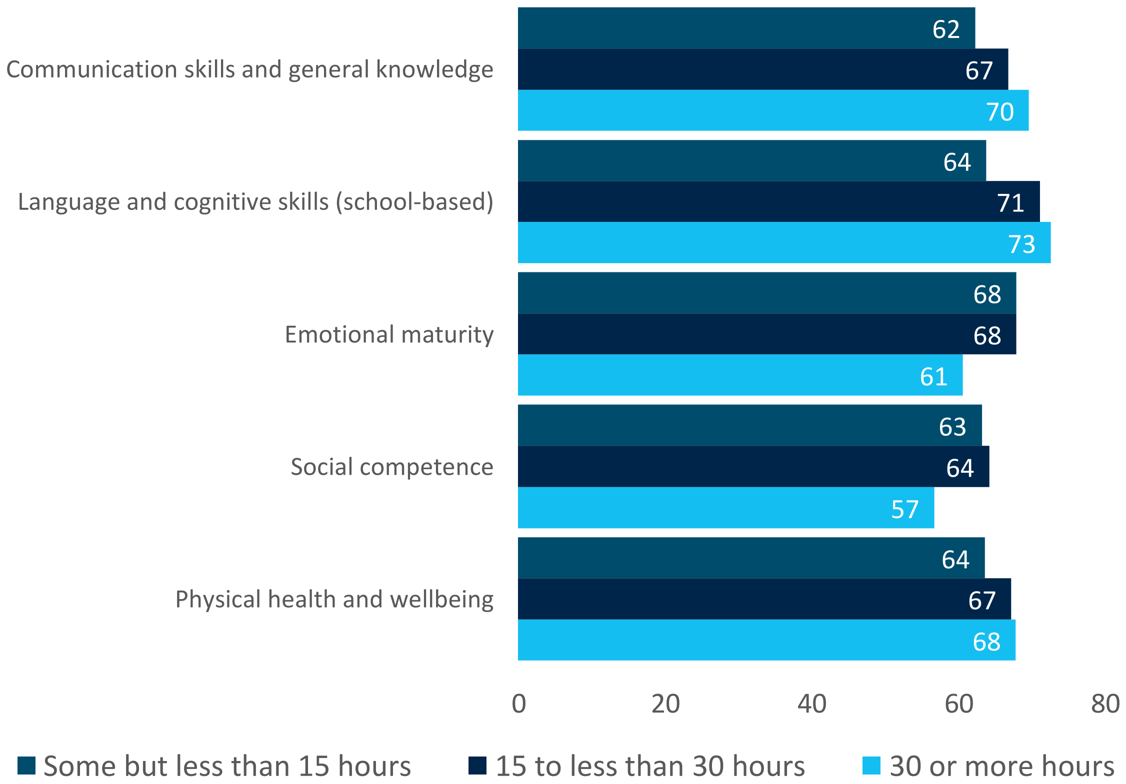
Overall, these equity groups had a lower rate of being developmentally on track on all domains regardless of average child care hours. Again, children who attended neither child care nor preschool had the lowest rate of being developmentally on track for each group.

The lowest SEIFA and single parent cohorts had a similar pattern to the entire AEDC cohort, with children who attended preschool without child care having the highest rates of being developmentally on track. Aboriginal and Torres Strait Islander and LBOTE children had a different trend, with children with 15 to less than 30 hours being the most likely to be developmentally on track on all domains, and those with over 30 hours having higher rates of being developmentally on track than those with less than 15 hours (including those with no formal child care hours).

Aboriginal and Torres Strait Islander children and the AEDC domains

For Aboriginal and Torres Strait Islander children using formal child care, rates of being developmentally on track on individual AEDC domains were generally higher for children in groups with higher average weekly charged hours. For the social competence and emotional maturity domains, however, Aboriginal and Torres Strait Islander children with 30 or more hours had lower rates of being developmentally on track than those with fewer hours.

Figure 8: Proportion (%) of Aboriginal and Torres Strait Islander children who were developmentally on track, by average weekly child care hours, for each domain, 2018 AEDC cohort.

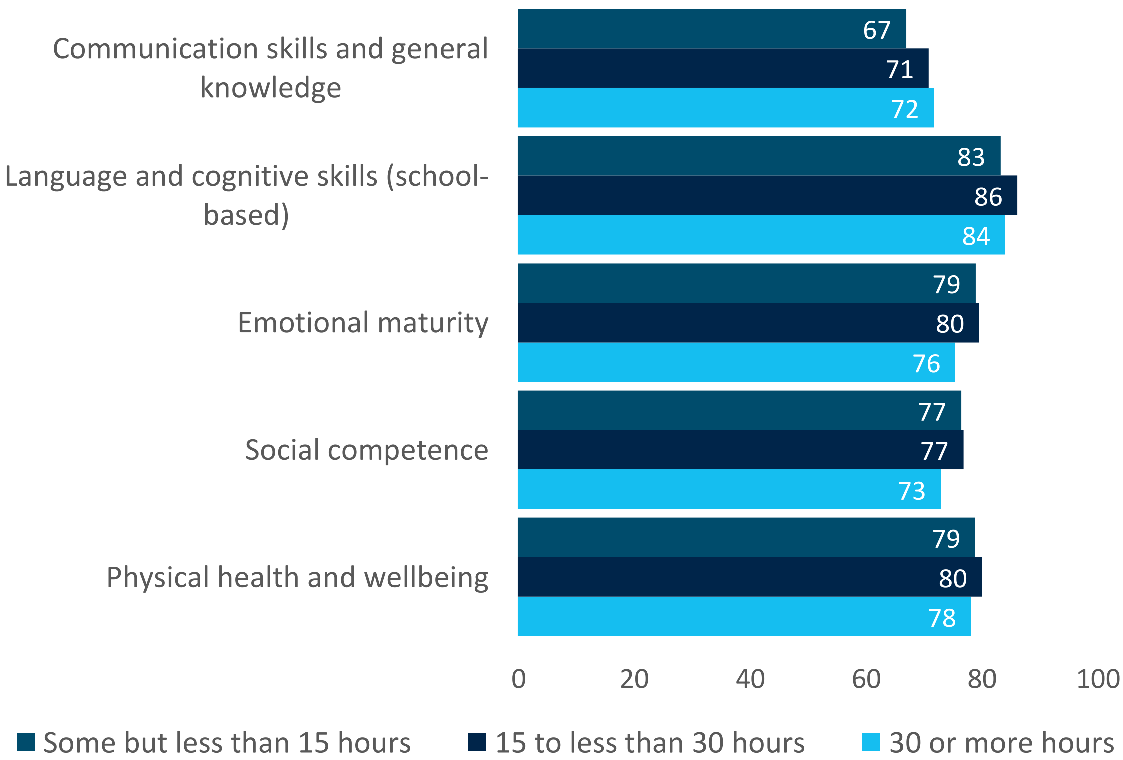


Source Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.

Children with a Language Background Other Than English (LBOTE) and the AEDC domains

For LBOTE children using formal child care, rates of being developmentally on track for the individual AEDC domains were broadly similar for children with different average charged child care hours. Those with 15 to less than 30 hours generally had the highest rate of being developmentally on track; with the exception being the communication skills and general knowledge domain where those with 30 or more hours were most likely to be developmentally on track.

Figure 9: Proportion (%) of LBOTE children who were developmentally on track, by average weekly child care hours, for each domain, 2018 AEDC cohort.



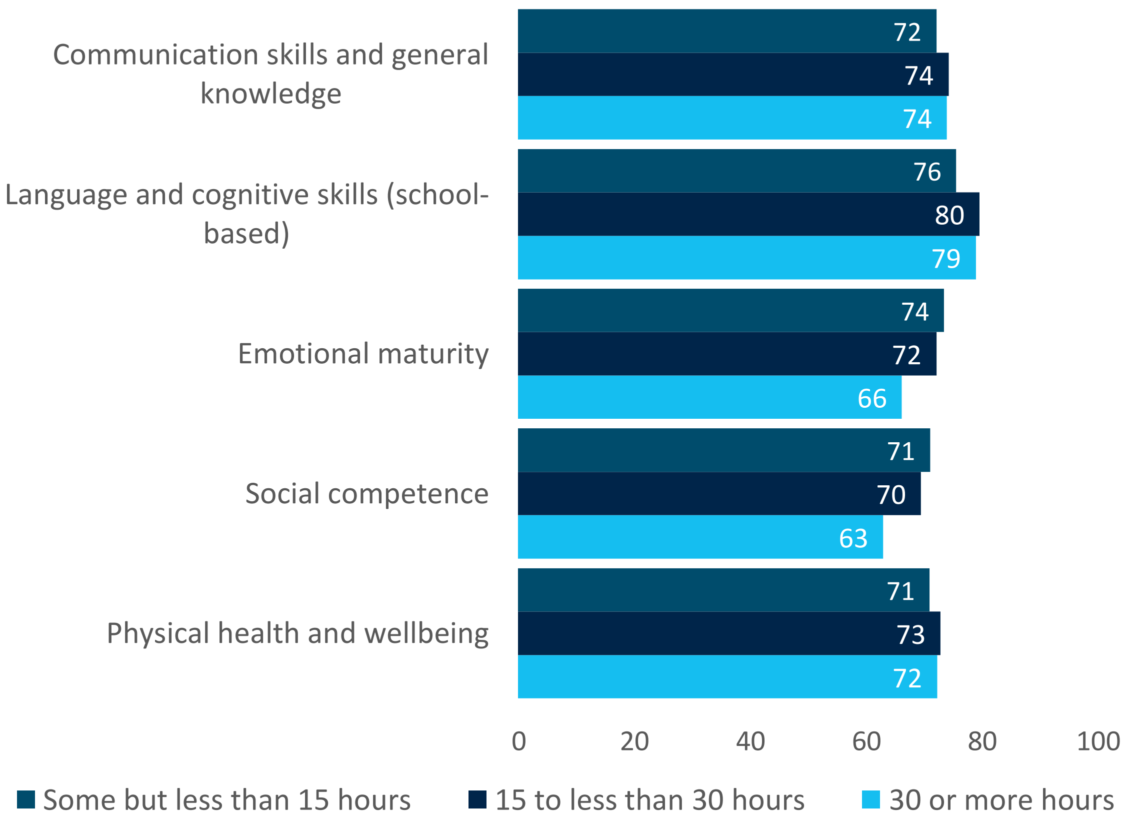
Source: Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.

Notes: Children were classified as LBOTE if at least one of their parents' language backgrounds was non-English.

Children from single parent households and the AEDC domains

For children from single parent households using formal child care, rates of being developmentally on track were lower on the emotional maturity and social competence domains for groups with higher average charged hours. For the other domains, children with 15 to less than 30 hours had the highest rate of being developmentally on track, though the differences between groups were smaller.

Figure 10: Proportion (%) of children from single parent households who were developmentally on track, by average weekly child care hours, for each domain, 2018 AEDC cohort.

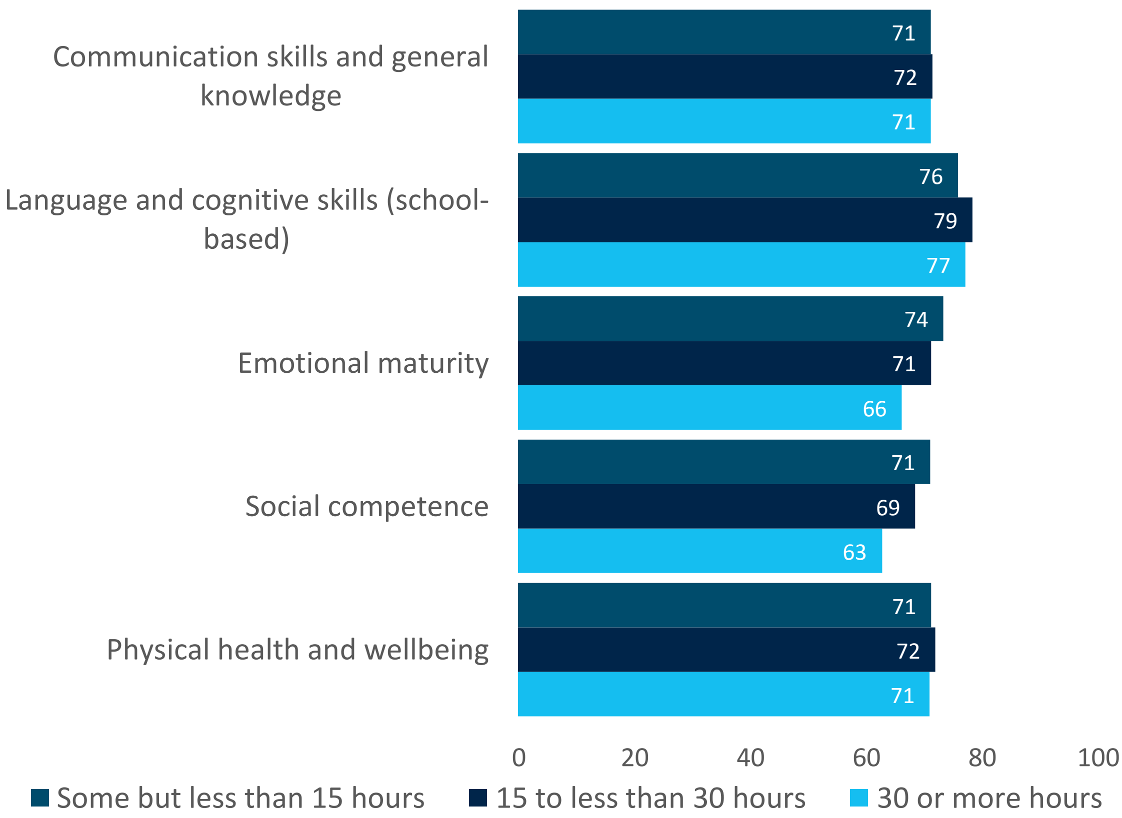


Source: Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.

Children from low socioeconomic backgrounds and the AEDC domains

Children from the lowest SEIFA quintile using formal child care displayed similar patterns of being developmentally on track across the domains to those from single parent households. Rates of being developmentally on track were lower on the emotional maturity and social competence domains for groups with higher average charged hours. For the other domains, children with 15 to less than 30 hours had the highest rate of being developmentally on track, though the differences between groups were smaller.

Figure 11: Proportion (%) of children in the lowest household SEIFA quintile who were developmentally on track, by average weekly child care hours for each domain, 2018 AEDC cohort.



Source: Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.

Notes: SEIFA quintile refers to the lowest Socio-Economic Index for Areas (SEIFA) index that the child had from birth to AEDC assessment, with 1 being the lowest SEIFA quintile.

References

Belsky J, Vandell DL, Burchinal M, Clarke-Stewart KA, McCartney K, Owen MT, and NICHD Early Child Care Research Network (2007) 'Are there long-term effects of early child care?', *Child development*, *78*(2), 681–701, doi: 10.1111/j.1467-8624.2007.01021.x.

DESE (Department of Education, Skills and Employment) (2021) *Universal Access National Partnership*, DESE website, accessed 22 December 2021.

Kalb G, Tabasso D and Zakirova R (2014) *'Children's participation in early childhood education and care, and their developmental outcomes by Year 5: A comparison between disadvantaged and advantaged children.'* Report for the Department of Education, Employment and Workplace Relations, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

NICHD Early Child Care Research Network (Ed.) (2005) *Child care and child development: Results from the NICHD study of early child care and youth development*. Guilford Press.

PC (Productivity Commission) (2024) A path to universal early childhood education and care: Inquiry report – volume 2, supporting papers. Report no. 106. 28 June 2024.

Australian Government
Department of Education

The First Five Years: What makes a difference?

3.4 Estimating the effect of child care on children's developmental outcomes

Key findings

The results presented in the other factsheets simply described the different outcomes for different groups of children. This basic analysis was extended by statistical techniques that allow better estimates of the effect of child care on children's developmental outcomes. These techniques aim to take account of the other differences in children's circumstances that are visible in the data, and so focus on outcomes due to different child care use rather than any other underlying differences between groups.

- As average weekly charged hours increased, the risk of developmental vulnerability on one or more domains (DV1) initially fell and then rose. This same pattern was observed for total hours before adjustment, but risk ratios increased with total hours after controlling for the different circumstances of different groups.
- Child care was generally a protective factor 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.
- The pattern of risk ratios for the physical health and wellbeing domain were more complex, with child care attendance a protective factor for some patterns of attendance and a risk factor for others. In general, risk ratios increased with increasing average hours and broadly fell with increasing total hours.
- Children's outcomes improved with increasing child care quality. Attending above standard child care reduced the child developmental vulnerability on one or more domains (DV1) and in each domain, compared to attending at standard child care. Not attending formal child care was a protective factor on some domains and a risk for others.

Moving from description to inference

In sections 3.1 to 3.3, we presented descriptive analysis which showed variations in children's developmental vulnerability by child care attendance, quality of child care and hours of child care. While this descriptive analysis shows trends between developmental vulnerability and child care use, it cannot tell whether there is a causal relationship. To establish a causal effect between child care attendance and developmental vulnerability, additional analysis is needed to account for other factors contributing to developmental vulnerabilities (Hernan and Robins 2020).

The *observable* socio-economic factors and health factors described in section 2 influence child development. For example, children who attended child care may come from higher income households than those who did not. As such measuring the influence of child care attendance is in part measuring the effect of household income rather than just the effect of child care. Appropriate statistical analysis techniques can be used to adjust for observable confounding factors and better estimate the effect of child care.

There are also *unobservable* factors that affect child development, such as quality of home care, parenting behaviour, and genetic factors. We are unable to adjust for these factors in our analysis, and thus cannot conclude causal effects for child care on developmental vulnerabilities.

Nonetheless, in this section, we aim to estimate the effect of child care on AEDC domain developmental vulnerabilities adjusting for imbalance in confounding covariates (Tartaglia and Knapp (unpublished)). Two approaches were used for the analysis: G-computation and Inverse Propensity Weighting (IPW) (see the *Methodology* section for details). These methods adjust for available confounding factors that predict child care use and developmental vulnerability. *Unobserved* variables, for example parenting style and quality of home care, cannot be taken into account. As such, the results establish a better estimate of the effect of child care on developmental vulnerability, but require further research to be confirmed.

Risk ratios

In this section, the risk ratio of a particular treatment (for example, using child care for a particular range of average weekly charged hours) is reported. The risk ratio is the ratio of the risk of being developmentally vulnerable for the treated group (those using child care in this way) compared to the risk of being developmentally vulnerable in the reference group (generally those not attending formal child care). A risk ratio of 1.2 for a particular treatment and domain means children with that treatment are 20 per cent more likely to be developmentally vulnerable in that domain than the reference group. A risk ratio of 0.9 means a child is 10 per cent less likely to be developmentally vulnerable. The risk ratio derived from associative analysis without adjusting for any confounding factors (for example, the results in section 3.1) is termed the unadjusted risk ratio. The adjusted risk ratio uses statistical techniques to adjust for underlying differences in characteristics between the groups.

The difference between the adjusted and unadjusted risk ratios is due to the confounding factors which are associated with both child care use and child development, such as parental education. Children who attend child care are more likely to experience socio-economic factors such as higher parental education and income, which are associated with lower rates of developmental vulnerability. This means that adjusted risk ratios for child care use are typically higher than unadjusted risk ratios – the unadjusted ratios also pick up some of these underlying advantages, while the adjusted ratios aim to focus on just the effect of child care usage.

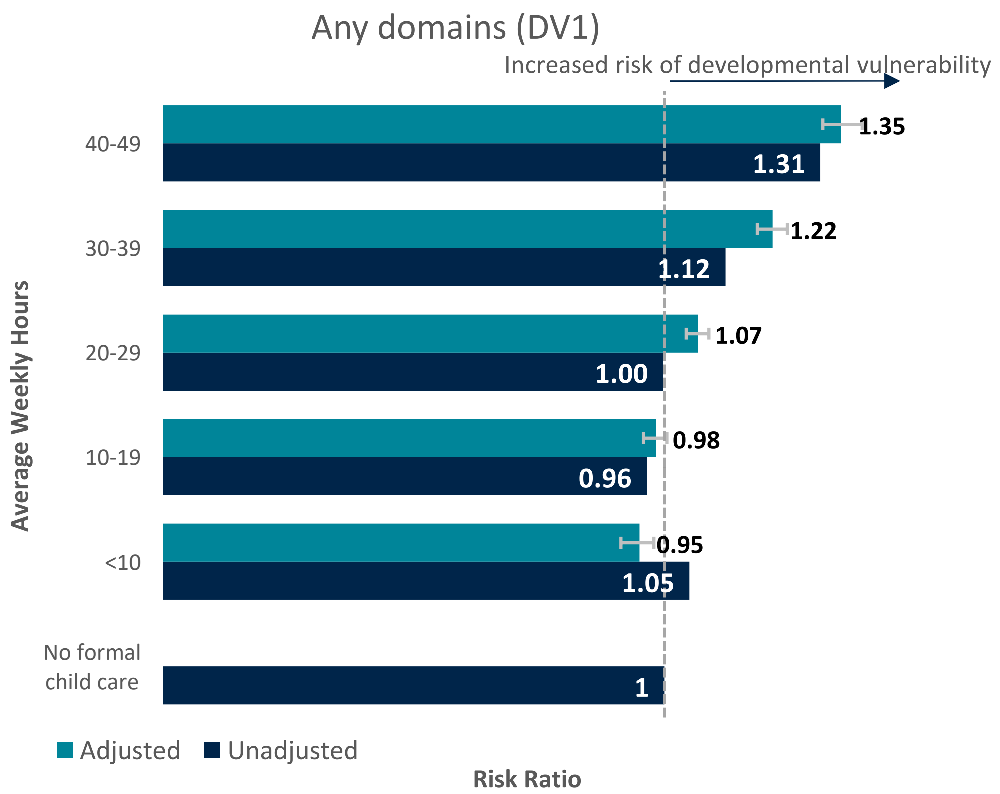
Child care attendance and hours

In section 3.1 it was shown that attending child care is associated with a lower rate of developmental vulnerability on one or more domains. Further analysis, in section 3.3, which extends the binary child care attendance to different attendance patterns using average weekly and total hours, showed large variations in children's developmental outcomes depending on patterns of attendance.

In this section, we use the statistical techniques (descripted in the *Methodology* paper) to estimate the effect of average weekly hours, total hours, and a combination of weekly intensity and overall hours in child care enrolment on children's risk of developmental vulnerability on one or more domains (DV1) and in each domain.

In Figure 1 and 2, the adjusted risk ratios for attending child care for different average weekly hours and total hours are shown.

Figure 1. Risk ratios for the effect of average weekly hours of enrolment in child care on developmental vulnerability on one or more domains compared to not attending formal child care.



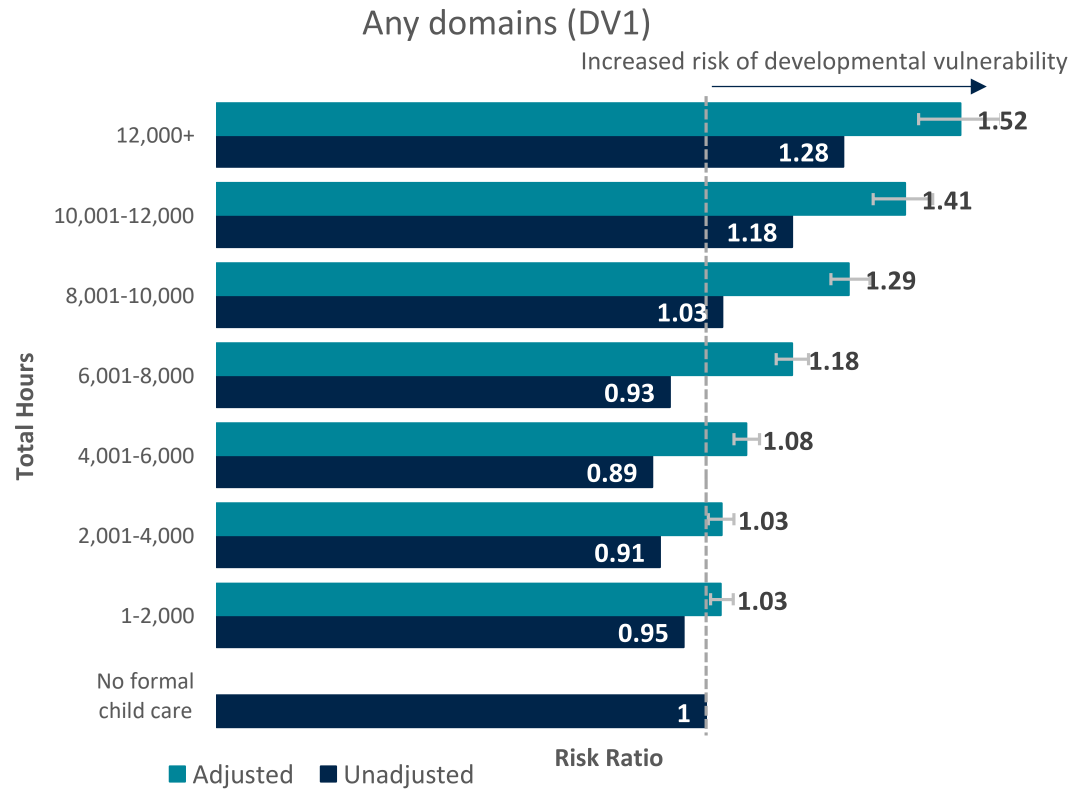
Source: Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.

Notes: This figure compares children from the 2018 cohort of the Australian Early Development Census. DV1: Developmentally vulnerable on one or more domains. N =232,693. For simplicity, results from G-computation were the only statistical inference results shown. For the <10 group, the unadjusted risk ratio is 1.05; after adjustment, the risk ratio reduced to 0.95.

Child care usage at less than 20 average weekly charged hours was a protective factor for developmental vulnerability on one or more domains. Before adjustment, the risk ratios of being developmentally vulnerable on one or more domains were 1.05 and 0.96 for the group with average hours of less than 10 and 20 respectively. The corresponding risk ratio changed to 0.95 and 0.98 after adjusting for covariates. Higher average weekly hours resulted in risk ratios above 1, indicating higher risks of developmental vulnerability compared to children who did not attend child care.

Total hours appeared to be a risk factor for DV1 across all levels (figure 2) after adjustment, though before adjustment the risk ratio was less than 1 for less than 8,000 total hours. As total hours are the sum of charged hours from birth to the year before school, there can be large variations in the attendance patterns within each group: for example, children who attended child care intensively when they were young and other children who attended child care only when they were older can be in the same group.

Figure 2. Risk ratios for the effect of Total enrolment hours of child care on developmental vulnerability on one or more domains compared to not attending formal child care.



Source: Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.

Notes: This figure compares children from the 2018 cohort of the Australian Early Development Census. DV1: Developmentally vulnerable on one or more domains. N =232,693. For simplicity, results from G-computation were the only statistical inference results shown.

To account for both the intensity and total amount of child care usage, we showed adjusted risk ratios for combinations of average weekly and total lifetime child care hours (Table 1). The horizontal bar at the bottom of the table summarises the risk ratios for average weekly hours (as already shown in figure 1), while the vertical bar to the right does the same for total hours (just like together with the effect of total hours (as shown in figure 2).

The results were less precisely estimated when the number of children in a given combination of average weekly and total hours were small. For this reason, cells with fewer than 200 children, or risk ratios with a confidence interval equal to or exceeding one, have been removed (the number of children in each combination of average and total hours is shown in Figure 4b of section 3.3). The small number of children with 50 or more average weekly hours are also removed, because at most a single such cell per table passes these conditions.

Overall, the results show that consistent use of lower weekly hours of child care was associated with lower rates of developmental vulnerability. This held when considering the risk of developmental vulnerability across any of the domains or in each domain separately.

Table 1. Adjusted risk ratios for the effect of combined child care average and total hours on developmental vulnerability in any AEDC domain compared to no formal child care attendance.

Any vulnerability (DV1), G-computation adjusted

Total Hours	Average Weekly Hours <10	10-19	20-29	30-39	40-49	
1-2,000	0.95	1.03	1.16	1.30		1.03
2,001-4,000		0.93	1.08	1.27	1.07	1.03
4,001-6,000		0.84	1.04	1.23	1.31	1.08
6,001-8,000			1.00	1.25	1.37	1.18
8,001-10,000				1.21	1.36	1.29
10,001-12,000					1.36	1.41
12,000+					1.46	1.52
	0.95	0.98	1.07	1.22	1.35	

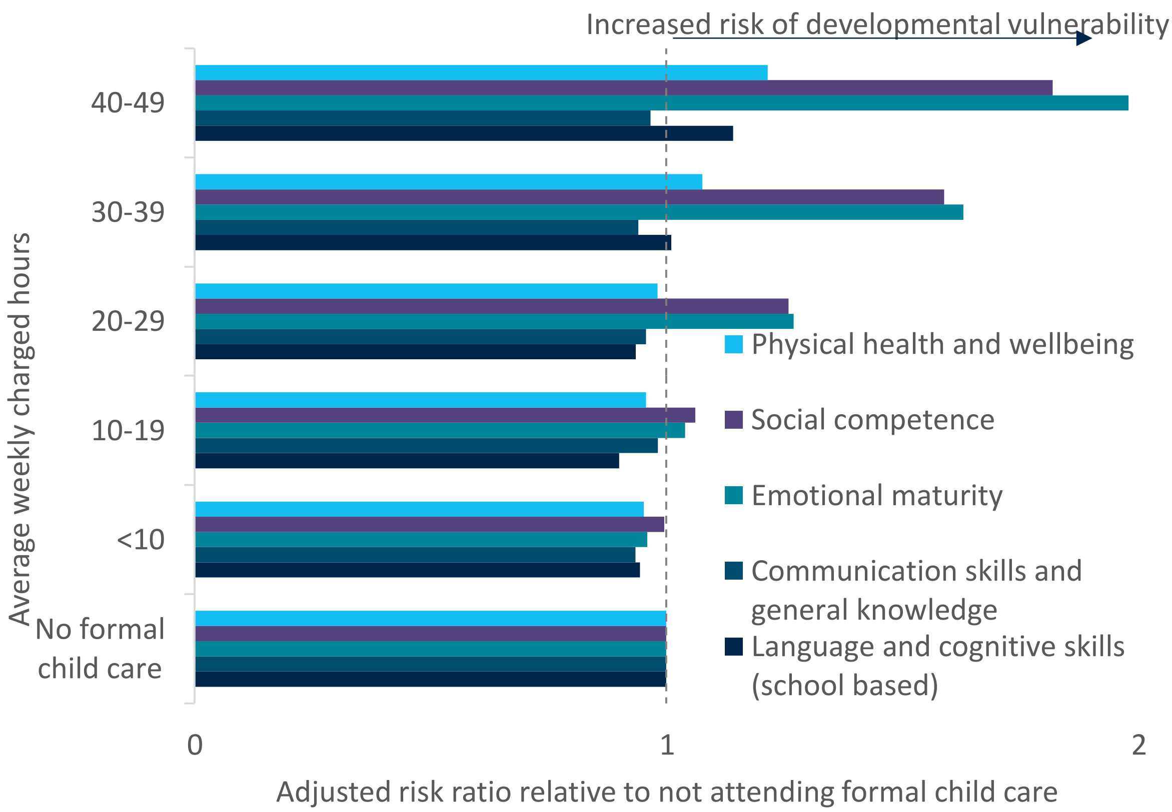
1.5
1
0.5

Source: Customised 'First Five Years' extract from the Multi-Agency Data Integration Project, accessed 2023.

Notes: This figure compares children from the 2018 cohort of the Australian Early Development Census. DV1: Developmentally vulnerable on one or more domains. N =232,693. For simplicity, results from G-computation were the only results shown. Reference is the group not attending formal child care (risk ratio=1). Only the central estimations were shown. Cells with fewer than 200 children or risk ratios with a confidence interval equal to or exceeding one have been removed.

Figure 3 summarises on a single chart the risk ratios by average weekly charged hours for the separate AEDC domains.

Figure 3. Adjusted risk ratios for the developmental vulnerability by average weekly charged hours and AEDC domain, compared to no formal child care attendance.



Source: Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.

Notes: Children may have separately attended preschool, which was not captured in the CCMS data.

There are clear differences in the patterns of risk ratios across the domains and average weekly hours. Risk ratios for the communication skills and general knowledge domain remain below 1 for all levels of average hours, indicating that child care is generally a protective factor for this domain. For the language and cognitive skills (school-based) domain, risk ratios initially fall as average weekly hours increase, suggesting that some child care hours result in a reduced risk of developmental vulnerability on this domain. Only for higher levels of average weekly hours for this domain do the risk ratios exceed 1, indicating an increased risk of developmental vulnerability. For the social competence and emotional maturity domains, risk ratios increase significantly as average weekly hours increase and reach high levels.

Cognitive Domains

Focusing on the different AEDC domains, formal child care use generally served as a protective factor for cognitive domains (communication skills and general knowledge and language and cognitive skills (school-based)), as shown in section 3.1, before any adjustment. The protective effect remained when considering average weekly hours of attendance (figure 6 of section 3.3), for the group with less than 30 average weekly charged hours.

The adjusted risk ratios for the effect of child care on language and cognitive skills (school-based), and the communication skills and general knowledge, are shown in Table 2 and Table 3 respectively.

For less than 30 average weekly charged hours, and between 2,000 to 8,000 total hours, the risk ratio of being developmentally vulnerable in the language and cognitive skills (school-based) domain were less than 1. Adjusted risk ratios for the language and cognitive skills (school-based) domain (Table 2) generally fell with total hours and increased with average hours.

Table 2. Adjusted risk ratios for the effect of combined child care average and total hours on developmental vulnerability in the Language and cognitive skills (school-based) AEDC domain compared to no formal child care attendance.

Language and cognitive skills (school-based), G-computation adjusted

Total Hours	Average Weekly Hours <10	10-19	20-29	30-39	40-49	
1-2,000	0.93	0.97	1.17	1.28		1.00
2,001-4,000		0.79	0.91	1.14	1.09	0.89
4,001-6,000		0.60	0.88	1.09	1.27	0.94
6,001-8,000			0.81	0.95	1.21	0.97
8,001-10,000				0.85	1.16	1.03
10,001-12,000					0.99	1.02
12,000+					1.04	1.18
	0.94	0.90	0.94	1.01	1.14	

1.5
1
0.5

Source: Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.

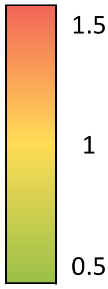
Notes: This figure compares children from the 2018 cohort of the Australian Early Development Census. N =232,693. For simplicity, results from G-computation were the only causal inference results shown. Further discussion available in methodology. Reference is the group not attending formal child care (risk ratio=1). Only the central estimations were shown. Cells with fewer than 200 children or risk ratios with a confidence interval equal to or exceeding one have been removed.

The communication skills and general knowledge domain exhibited the same general pattern as the language and cognitive skills (school-based) domain, with risk ratios generally falling with total hours and increasing with average hours. Child care attendance remained a protective factor even up to 50 average weekly charged hours. Risk ratios were only above 1, indicating higher risks of developmental vulnerability, for children with less than 4,000 total hours or 30 or more average weekly charged hours.

Table 3. Adjusted risk ratios for the effect of combined child care average and total hours on developmental vulnerability in the Communication skills and general knowledge AEDC domain compared to no formal child care attendance.

Communication skills and general knowledge, G-computation adjusted

Total Hours	Average Weekly Hours <10	10-19	20-29	30-39	40-49	
1-2,000	0.92	1.05	1.18	1.18		1.04
2,001-4,000		0.88	1.01	1.17	1.03	0.98
4,001-6,000		0.67	0.83	0.97	0.99	0.87
6,001-8,000			0.77	0.87	1.08	0.89
8,001-10,000				0.80	0.94	0.88
10,001-12,000					0.86	0.89
12,000+					0.91	0.90
	0.93	0.98	0.96	0.94	0.97	



Source: Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.

Notes: This figure compares children from the 2018 cohort of the Australian Early Development Census. N =232,693. For simplicity, results from G-computation were the only causal inference results shown. Further discussion available in methodology. Reference is the group not attending formal child care (risk ratio=1). Only the central estimations were shown. Cells with fewer than 200 children or risk ratios with a confidence interval equal to or exceeding one have been removed.

Non-Cognitive Domains

Non-cognitive domains showed a different pattern to the cognitive domains. Formal child care attendance appeared to be a risk factor for the emotional maturity and social competence domains before any adjustment, as shown in section 3.1. Higher average weekly hours were associated with significantly lower rates of being developmentally on track for these two domains (figure 6 of section 3.3).

The adjusted risk ratios for the effect of child care on the emotional maturity and social competence domains are shown in Table 4 and Table 5 respectively.

As with the descriptive analysis, the adjusted ratio shows that higher average weekly or total hours of child care use were associated with elevated risk ratios on the emotional maturity domain (Table 4). This effect was particularly pronounced for high average weekly hours.

Table 4. Adjusted risk ratios for the effect of combined child care average and total hours on developmental vulnerability in the Emotional maturity AEDC domain compared to no formal child care attendance.

Emotional maturity, G-computation adjusted

Total Hours	Average Weekly Hours <10	10-19	20-29	30-39	40-49	
1-2,000	0.96	1.06	1.25	1.22		0.89
2,001-4,000		1.01	1.18	1.46	1.49	0.99
4,001-6,000		1.07	1.35	1.60	2.00	1.19
6,001-8,000			1.34	1.65	1.81	1.29
8,001-10,000				1.79	2.03	1.56
10,001-12,000				1.37	2.04	1.79
12,000+					2.38	2.18
	0.98	1.02	1.20	1.50	1.85	

Source: Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.

Notes: This figure compares children from the 2018 cohort of the Australian Early Development Census. N =232,693. For simplicity, results from G-computation were the only causal inference results shown. Further discussion available in methodology. Reference is the group not attending formal child care (risk ratio=1). Only the central estimations were shown. Cells with fewer than 200 children or risk ratios with a confidence interval equal to or exceeding one have been removed.

Table 5. Adjusted risk ratios for the effect of combined child care average and total hours on developmental vulnerability in the Social competence AEDC domain compared to no formal child care attendance.

Social competence, G-computation adjusted

Total Hours	Average Weekly Hours <10	10-19	20-29	30-39	40-49	
1-2,000	0.99	1.10	1.25	1.29		1.10
2,001-4,000		1.02	1.24	1.55	1.40	1.18
4,001-6,000		0.92	1.27	1.56	1.79	1.35
6,001-8,000			1.29	1.62	1.75	1.57
8,001-10,000				1.63	1.85	1.77
10,001-12,000					1.87	1.96
12,000+					1.99	2.23
	1.00	1.06	1.26	1.59	1.82	

Source: Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.

Notes: This figure compares children from the 2018 cohort of the Australian Early Development Census. N =232,693. For simplicity, results from G-computation were the only causal inference results shown. Further discussion available in methodology. Reference is the group not attending formal child care (risk ratio=1). Only the central estimations were shown. Cells with fewer than 200 children or risk ratios with a confidence interval equal to or exceeding one have been removed.

The social competence domain also showed elevated risk ratios with increasing weekly child care hours and increasing total hours (Table 5).

Table 6. Adjusted risk ratios for the effect of combined child care average and total hours on developmental vulnerability in the Physical health and wellbeing AEDC domain compared to no formal child care attendance.

Physical health and wellbeing, G-computation adjusted

Total Hours	Average Weekly Hours <10	10-19	20-29	30-39	40-49	
1-2,000	0.95	1.02	1.11	1.36		1.02
2,001-4,000		0.87	1.02	1.25	0.88	0.98
4,001-6,000		0.76	0.92	1.12	1.38	0.98
6,001-8,000			0.82	1.01	1.19	1.00
8,001-10,000				1.00	1.27	1.14
10,001-12,000				0.81	1.14	1.22
12,000+					1.11	1.19
	0.95	0.96	0.98	1.08	1.21	

Source: Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.

Notes: This figure compares children from the 2018 cohort of the Australian Early Development Census. N =232,693. For simplicity, results from G-computation were the only causal inference results shown. Further discussion available in methodology. Reference is the group not attending formal child care (risk ratio=1). Only the central estimations were shown. Cells with fewer than 200 children or risk ratios with a confidence interval equal to or exceeding one have been removed.

Risk ratios for the physical health and wellbeing domain were more complex, with formal child care a protective factor for some patterns of attendance and a risk factor for others. In general, risk ratios increased with increasing average hours and broadly fell with increasing total hours. Taking average weekly and total hours separately, child care was estimated to be a protective factor for this domain for less than 30 average charged weekly hours, or between 2,000 and 6,000 total hours.

Child care quality

The descriptive results in section 3.2 compare rates of developmental vulnerability for children who attend child care at different quality ratings. Children who attend child care have lower rates of developmental vulnerability than children who do not, except for those attending child care with a not yet at standard quality rating. As has been shown from Australian national data, there are inequities to accessing high quality care associated with geographic and socio-economic disadvantage (Tang et al. 2024).

In this section, we use statistical techniques to estimate the effect of different child care quality (not yet at standard child care, above standard child care, and no formal child care, with at standard child care as a reference group), to better understand the relationship between child care quality on children's risk of developmental vulnerability.

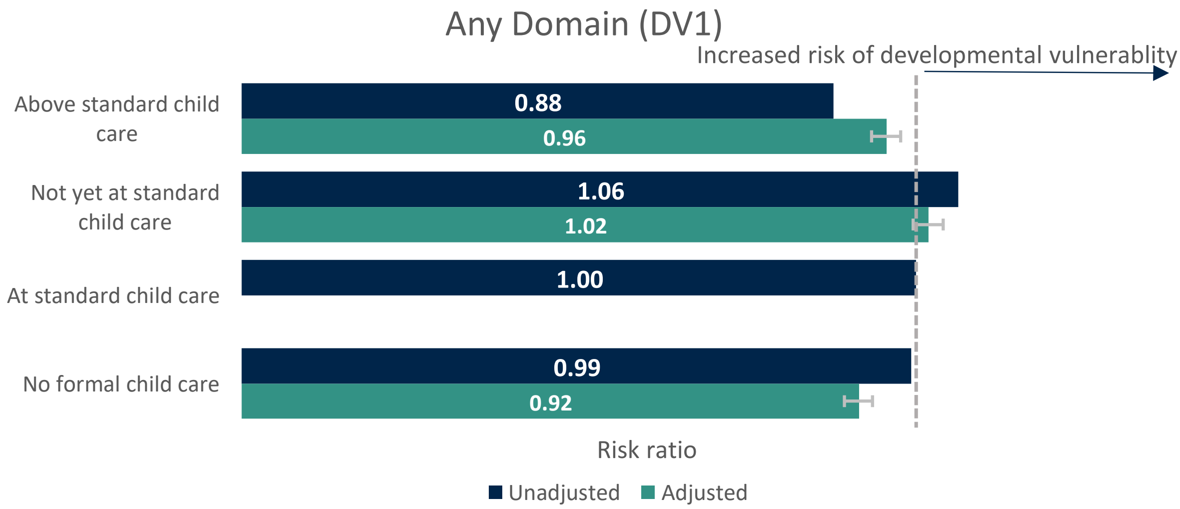
We checked the distribution of child care hours for the different quality group to make sure that differences in outcomes were not being driven by differences in hours across the quality groups. The distributions of average weekly child care hours are similar for the three groups using formal child care. The group who did not use formal child care is also included, noting its diversity (and in particular, the different outcomes for those who did and did not attend preschool).

This section focusses on the effect of quality on outcomes. Other analysis that considered the combined effects of quality and hours suggests that these dimensions can be considered separately, rather than there being significant interaction effects.

Overall, the adjusted risk ratios for DV1 and each domain showed the higher the quality rating received by the child care provider, the better the outcome for the child. Attending above standard child care decreases the risk of being developmentally vulnerable in DV1 and in each domain when compared with attending at standard child care. This is consistent with finding from similar studies (focusing on different quality areas) in Australia (Rankin et al. 2024), where it is shown that children in services not yet at or at standard had higher rates of developmental vulnerability compared to those in services rated as above standard.

Figure 4 shows risk ratios for being developmentally vulnerable in one or more domains (DV1). Compared with attending at standard child care, attending above standard quality child care decreases the risk of being developmentally vulnerable (the adjusted risk ratio is 0.96 – that is, the risk of vulnerability is 4 per cent lower). Attending not yet at standard quality child care, on the other hand, increases the risk of being developmentally vulnerable by 2 per cent. The risk ratio for the group not attending formal child care is comparable with those attending At standard child care before adjustment, and shows an 8 percentage point lower risk of being developmentally vulnerable.

Figure 4. Risk ratios for the effect of child care quality on developmental vulnerability on one or more domains compared to attending *At standard child care*.



Source: Custom Multi-Agency Data Integration Project (MADIP) extract linked to Australian Early Development Census (AEDC) 2018 and Child Care Management System (CCMS).

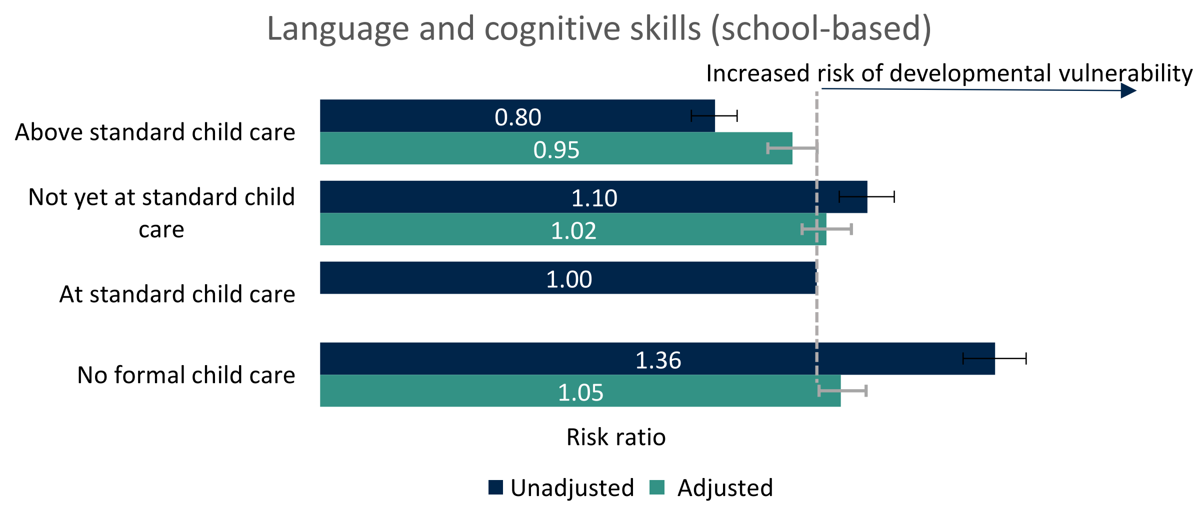
Notes: This figure compares children from the 2018 cohort of the Australian Early Development Census. Error bars are 95 per cent confidence interval. (N=53,795 not attending formal child care; 41,873 attending above standard quality; 53,368 attending at standard quality; 48,566 attending not yet at standard quality; 35,106 attended child care with unknown quality were removed from this analysis). For simplicity, results from G-computation were the only statistical inference results shown. Further discussion available in methodology. Children may have separately attended preschool, which was not captured in the CCMS data.

Cognitive Domains

The pattern of results for different quality levels on cognitive domains showed that attendance across the quality ratings (compared with not attending formal child care) generally has a protective effect.

In the Language and cognitive skills (school-based) domain (Figure 5), attending Above standard child care lowered the risk of being developmentally vulnerable by 5 per cent. Both attending not yet at standard child care and no formal child care increased the risk of being developmentally vulnerable.

Figure 5. Risk ratios for the effect of child care quality on developmental vulnerability in the Language and cognitive skills (school-based) domain, compared to attending *At standard child care*.

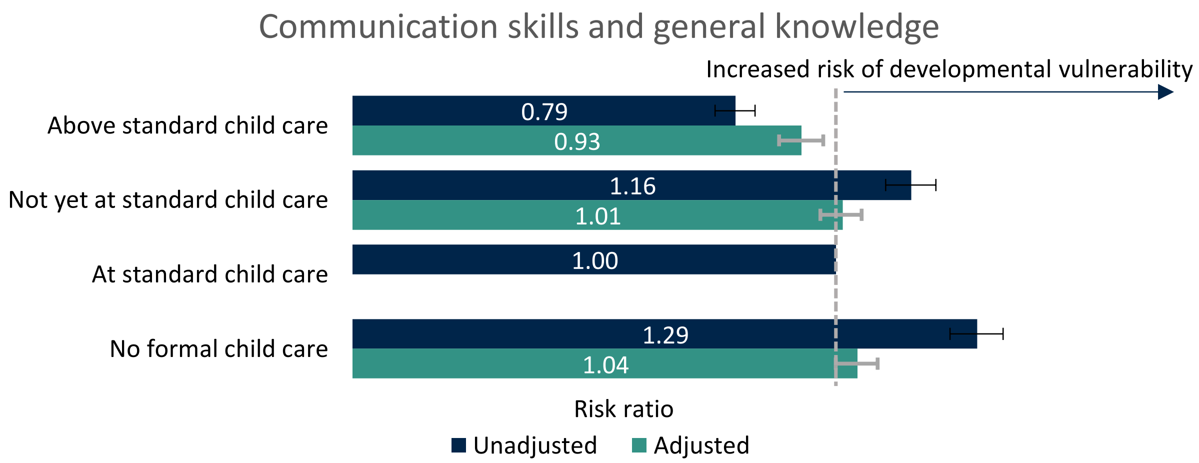


Source: Customised ‘First Five Years’ extract from the Multi-Agency Data Integration Project.

Notes: This figure compares children from the 2018 cohort of the Australian Early Development Census. Error bars are 95 per cent confidence intervals. N =232,693. For simplicity, results from G-computation were the only statistical inference results shown. Further discussion available in methodology. Children may have separately attended preschool, which was not captured in the CCMS data.

Attending above standard quality child care was a protective factor for developmental vulnerabilities both before and after adjustment. Adjusting for the different characteristics of the groups lowers the apparent risk ratios of groups using not yet at standard or no formal child care, though the point estimates still show higher risks of being developmentally vulnerable than for children in at standard quality care.

Figure 6. Risk ratios for the effect of child care quality on developmental vulnerability in the Communication skills and general knowledge domain, compared to attending *At standard child care*.



Source: Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.

Notes: This figure compares children from the 2018 cohort of the Australian Early Development Census. Error bars are 95 per cent confidence intervals. N =232,693. For simplicity, results from G-computation were the only statistical inference results shown. Further discussion available in methodology. Children may have separately attended preschool, which was not captured in the CCMS data.

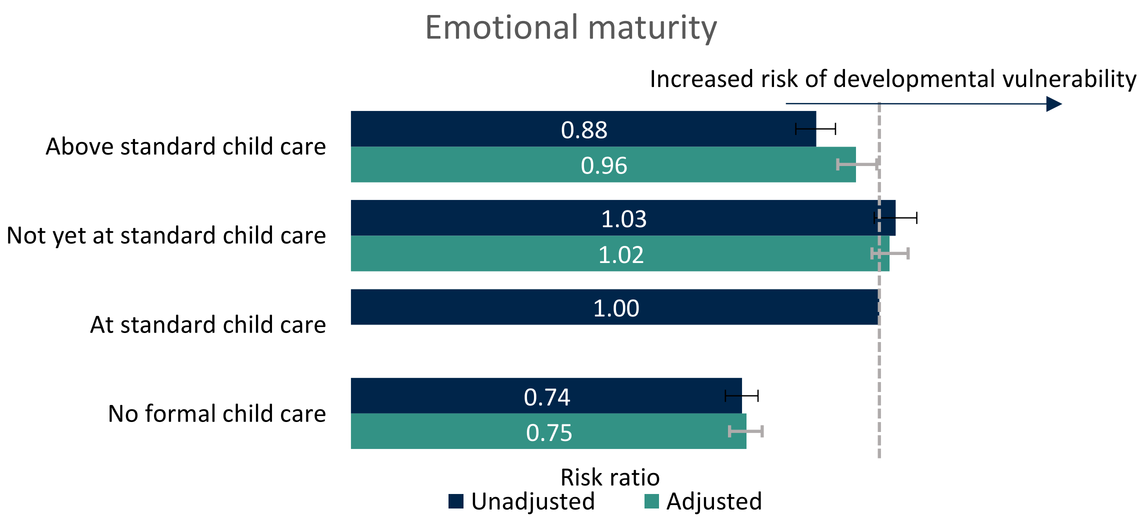
The same pattern was present for the communication skills and general knowledge domain. Above standard child care was estimated to be a statistically significant protective factor after adjustment, compared with those attending At standard child care.

Non-Cognitive Domains

Results for social competence and emotional maturity domains were broadly similar, while the physical health and wellbeing domain displayed a different pattern.

Adjusting for the different characteristics of the different groups had a relatively smaller effect for the emotional maturity domain (figure 7). Above standard child care was estimated to have a lower risk of developmental vulnerability on the emotional maturity domain compared to at standard child care. The risk ratio for the group not attending formal child care was the lowest, both before and after adjustment.

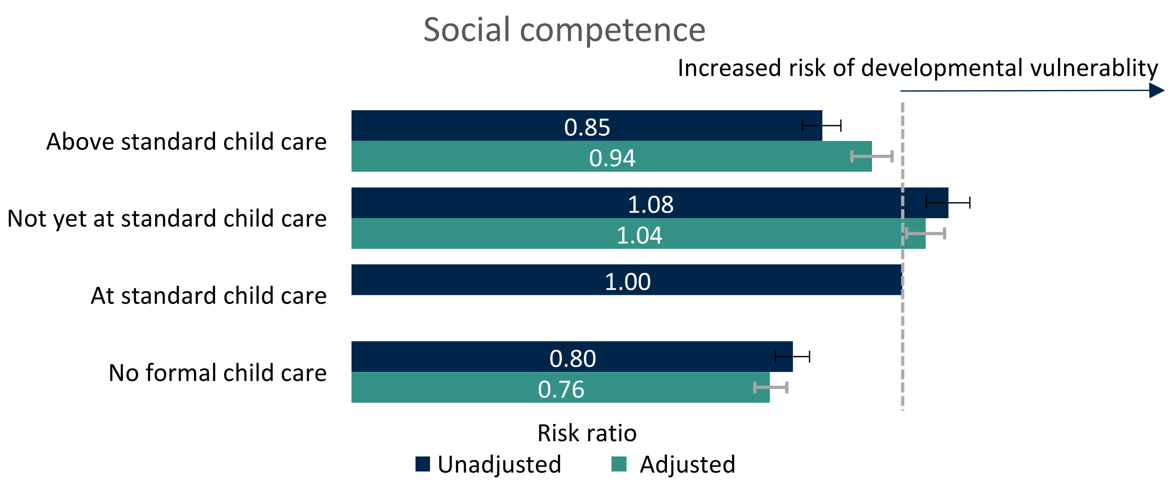
Figure 7. Risk ratios for the effect of child care quality on developmental vulnerability in the Emotional maturity domain, compared to attending *At standard child care*.



Source: Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.

Notes: This figure compares children from the 2018 cohort of the Australian Early Development Census. Error bars are 95 per cent confidence intervals. N =232,693. For simplicity, results from G-computation were the only statistical inference results shown. Further discussion available in methodology. Children may have separately attended preschool, which was not captured in the CCMS data.

Figure 8. Risk ratios for the effect of child care quality on developmental vulnerability in the Social competence domain, compared to attending *At standard child care*.



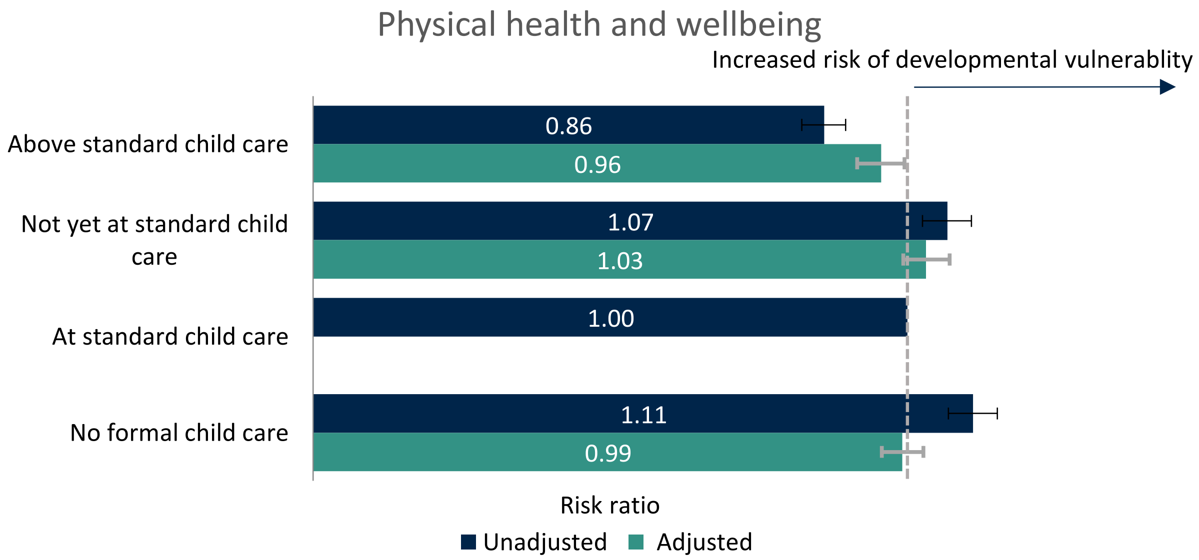
Source: Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.

Notes: This figure compares children from the 2018 cohort of the Australian Early Development Census. Error bars are 95 per cent confidence intervals. N =232,693. For simplicity, results from G-computation were the only statistical inference results shown. Further discussion available in methodology. Children may have separately attended preschool, which was not captured in the CCMS data.

Adjusting for the different characteristics of the groups accessing different quality child care produced a broadly similar pattern to the unadjusted results for the social competence domain

(Figure 8). Central estimates showed higher quality child care had lower risk ratios for developmental vulnerabilities on this domain, with a statistically significant difference between above and not yet at standard care. The group not attending formal child care had the lowest risk ratio, recognising the majority of these children attended preschool.

Figure 9. Risk ratios for the effect of child care quality on developmental vulnerability in the Physical health and wellbeing domain, compared to attending *At standard child care*.



Source: Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.

Notes: This figure compares children from the 2018 cohort of the Australian Early Development Census. Error bars are 95 per cent confidence intervals. N =232,693. For simplicity, results from G-computation were the only statistical inference results shown. Further discussion available in methodology. Children may have separately attended preschool, which was not captured in the CCMS data.

For the physical health and wellbeing domain (figure 9), the differences in the risk ratio for different groups are smaller after adjusting for the different characteristics of the cohorts. Central estimates were lower for higher quality child care, ranging from being a slight protective factor through to a slight risk factor, though results were generally not statistically significant.

References

Hernan M and Robins J (2020) 'Causal Inference: What If', *Boca Raton: Chapman & Hall/CRC*.

Rankin P, Staton S, Jones A, Potia AH, Houen S, Healey B and Thorpe K (2024) 'Linking quality and child development in early childhood education and care: Technical report', *Australian Education Research Organisation (AERO)*.

Tang A, Rankin P, Staton S and Thorpe K (2024) 'Access to high-quality early care and education: Analysis of Australia's national integrated data', *Early Childhood Research Quarterly* 67:352, doi:/10.1016/j.ecresq.2024.02.001.

Tartaglia E and Knapp S (unpublished) ‘Causal Inference in Education Data’, *Australian Government Department of Education, Skills and Employment (DESE)*.

Australian Government
Department of Education

The First Five Years: what makes a difference?

4. Conclusion

Summary

The First Five Years was developed as a flagship project under the Data Integration Partnership for Australia (DIPA) as a collaborative endeavour involving project partners from across government and the university sector. For the first time, administrative data on child care participation and quality was linked to a measure of childhood development, along with a broad range of family, social, economic and health data.

The project demonstrated the significant value of integrated data to contribute to the policy evidence base, focussing on two broad questions.

- What child and family characteristics contribute to children being developmentally vulnerable at age 4-6 years?
- How does child care attendance affect child developmental vulnerability, and do these effects differ based on child care usage patterns and quality of services?

The project used descriptive analysis, predictive modelling, and statistical inference techniques to explore associations between children's circumstances and child care use and their developmental outcomes in the first year of school.

Key results

The analysis found a range of child, family and social characteristics were associated with developmental vulnerability, recognising that developmental vulnerability is linked to the socio-economic conditions that families are experiencing rather than a direct result of factors such as a family's country of origin.

- Higher parent or carer educational attainment, higher neighbourhood socio-economic status and higher household income were associated with lower rates of developmental vulnerability.

- Children from language backgrounds other than English and those with a parent not born in an OECD country had higher rates of developmental vulnerability, as did children of mothers aged under 20 at the child's birth, children of single parents and children with parents experiencing mental ill-health.

Analysis of associations between child care and developmental vulnerability showed differences in rates of developmental vulnerability for different usage patterns and quality.

- 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.

Estimated effects of child care duration and quality on developmental vulnerability after controlling for differences in children's circumstances showed both positive and negative effects on different aspects of development.

- 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.

Limitations

The First Five Years project is a valuable addition to the policy evidence base, linking for the first time the developmental outcomes of an entire cohort of Australian children starting school to their child care use and family circumstances. At the same time, the results must be interpreted in line with the project's limitations.

More work is needed to infer causal relationships: while statistical techniques were used to adjust for differences in children's circumstances to better estimate the effect of child care on developmental vulnerability, the observational nature of the data and the inability to adjust for key factors (such as quality of home care) precludes drawing causal conclusions.

Some factors such as lower SES status, Aboriginal and Torres Strait Islander background, having mental ill-health, single parent household status and mothers under 20 are found to be predictors of children's developmental vulnerability. The association of family characteristics with developmental vulnerability may be due to the challenges and systemic barriers those families face, and this could be explored further.

The project has known data gaps, such as the lack of detailed information on government-run pre-school attendance. Other family and child characteristics known to affect a child's development, such as the quality of home care and exposure to family and domestic violence, are not available in the current study. Analysis results are also influenced by linkage quality: the group of children who were unable to be linked across datasets, and hence were excluded from the analysis, were more likely to be developmentally vulnerable in all domains. Thus, the analytical cohort is biased towards those who are less likely to be developmentally vulnerable.

Results were based on children's developmental outcomes as measured in the first year of school. The significance of these findings would be enhanced by further work looking at how these outcomes persist or not through the school years and beyond.

The project focussed on children's developmental outcomes assessed through the AEDC and did not consider broader benefits from child care use such as parental labour force participation and related social and economic gains.

This project reports only on child care services across three broad categories of quality, and does not engage with the dedicated service models that might better support development for different cohorts of children. The data and analytical findings did not report on factors such as the cultural safety or responsiveness of child care for Aboriginal and Torres Strait Islander children, especially the important of Social and Emotional Wellbeing (SEWB) which describes the social, emotional, spiritual and cultural wellbeing of Aboriginal and Torres Strait Islander people. Aboriginal Community-Controlled Organisations (ACCOs) have a special role in supporting the SEWB of Aboriginal and Torres Strait Islander children; however, some of them might not be included in the CCMS data (as they can be supported under the former Budget Based Funded services which are not included in the CCMS).

The results focus on formal early childhood education and care and kinship care is absent from the data, noting these arrangements can be important for Aboriginal and Torres Strait Islander children and people from certain other cultural backgrounds.

The results alone are not suitable for determining "optimal" amounts of child care for particular families: while the analysis can provide useful information on general benefits and risks, families will still need to consider these alongside their specific circumstances and preferences.

Next steps

The First Five Years project aimed to create an enduring data asset linking a measure of childhood development with family, social, economic and health data and data about child care centre attendance and quality.

The existing results do not exhaust the possibilities of the integrated data. Work on related questions remains active, including efforts to address some of the limitations described above. Another key priority is incorporating more recent data to provide more current results: administrative data covering the Child Care Subsidy system (which was introduced in 2018) is now available, along with later waves of the Australian Early Development Census.

Further analysis will build on the valuable contribution the First Five Years project has already made to the evidence base in this critical policy area.

The First Five Years: What makes a difference?

5.1 Methodology

This section details the data and methodology we employed to deliver the findings across the First Five Years project. This includes datasets, key analytical decisions, constructed factors and statistical techniques.

Contents

Data Linkage and Assembly

This project linked what was formerly known as Multi-Agency Data Integration Project (MADIP) data with Australian Early Development Census (AEDC) and Child Care Management System (CCMS) data.

MADIP has since been renamed the Person Level Integrated Data Asset (PLIDA). It is hosted by the Australian Bureau of Statistics (ABS) and is a partnership among Australian Government agencies to develop a secure and enduring approach for combining information on healthcare, education, government payments, personal income tax, and population demographics (including the Census) to create a comprehensive picture of Australia over time (ABS 2021a). This project used a customised MADIP extract that contained data from multiple government datasets (Table 1).

Table 1. First Five Years datasets

Dataset	Abbreviation	Source
Data Over Multiple Individual Occurrences	DOMINO	Department of Social Services
Data Exchange	DEX	Department of Social Services
Medicare Benefits Schedule	MBS	Department of Health
Medicare Enrolments Database	MEDB	Department of Health
Pharmaceutical Benefits Scheme	PBS	Department of Health
Census of Population and Housing		Australian Bureau of Statistics
Personal Income Tax	PIT	Australian Tax Office
Pay As You Go summaries	PAYG	Australian Tax Office
Registries Death Data		Australian Bureau of Statistics
Child Care Management System	CCMS	Department of Education
Australian Early Development Census	AEDC	Department of Education
National Quality Standard	NQS	Australian Children's Education and Care Quality Authority

The ABS provided the infrastructure and support to link these datasets and provided access to de-identified unit record files. Data is made available for analysis through the secure ABS DataLab environment (ABS 2021b).

We measured child developmental vulnerability using the AEDC. We measured child and family attributes that impact development using all available datasets.

Two cohorts were provided for analysis, 2015 AEDC and 2018 AEDC. These datasets were linked to the 2011 Census and 2016 Census respectively. For the First Five Years project we have published analysis on the 2018 cohort. The 2018 cohort was chosen as it is more recent, contains better quality data linkage and is closer to the respective Census. Data and methods described in this paper refer to analyses conducted with the 2018 AEDC cohort.

A custom child-carer relationship dataset was assembled that identified the child and their carer. This dataset identified the type of caring relationship (e.g., parent, grandparent, guardians) and the length of the relationship. The relationship scope dataset for the 2018 cohort used three data sources: Census 2016, DOMINO and CCMS. This dataset linked each child and carer to available MADIP data.

MADIP data was assembled to combine demographic data for individuals and key items including gender, birth, and death information. This combined demographic dataset for the 2018 cohort used five main data sources: MEDB, DOMINO, PIT, Census 2016, and Death Registrations. MADIP data was also assembled to combine location data for individuals. This combined location dataset identified residential location of individuals and the period they resided at each location. The location dataset for the 2018 cohort used four data sources: MEDB, DOMINO, PIT and Census 2016.

Developmental vulnerability

We explored developmental vulnerability in children using the five domains available on the AEDC (AEDC 2019a). The AEDC is nationwide data collection of early childhood development at the time children commence their first year of full-time school. The census involves teachers of children in their first year of full-time school completing a research tool, the Australian version of the Early Development Instrument. The Instrument collects data relating to five key areas of early childhood development domains. The domains and their descriptions are listed below (AEDC 2022):

- **Physical health and wellbeing** - Children's physical readiness for the school day, physical independence and gross and fine motor skills
- **Social competence** - Children's overall social competence, responsibility and respect, approach to learning and readiness to explore new things
- **Emotional maturity** - Children's pro-social and helping behaviours and absence of anxious and fearful behaviour, aggressive behaviour and hyperactivity and inattention
- **Language and cognitive skills (school-based)** - Children's basic literacy, advanced literacy, basic numeracy, and interest in literacy, numeracy and memory
- **Communication skills and general knowledge** - Children's communication skills and general knowledge based on broad developmental competencies and skills

The domain scores were originally developed by the Canadian Early Development Index. An Australian version was developed under license to form the AEDC (AEDC 2019b).

Primarily, analyses focused on the outcome of children being developmentally vulnerable on one or more domains (DV1). Some analyses also investigated developmental vulnerability in individual domains. We calculated the proportion of developmental vulnerability based on children who had a valid indicator. Individual indicators required a valid score in that domain exclusively. In the AEDC, the following criteria need to be met for a valid DV1 score:

- the child had at least five valid AEDC domain scores, or
- the child had one, two, three or four valid AEDC domain scores where at least one of these scores was categorised as developmentally vulnerable.

The proportion of children who were developmentally vulnerable on one or more domains was calculated by taking the number of children with at least one valid AEDC domain score which was categorised as developmentally vulnerable, divided by the number of those with a valid DV1 score (AEDC 2019c).

Another indicator, which focused on children who were developmentally on track on all domains, was also used in this project. 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.

The definition of being developmentally on track on all domains used in this work has a slightly different definition from both the AEDC OT5 indicator (AEDC 2022) and the Productivity commission's (PC) definition for the Closing the Gap (CtG) dashboard. In 2021 the AEDC introduced a strength-based indicator: OT5 (developmentally on track on all FIVE AEDC domains) and the corresponding OT5flag which indicates whether a child qualifies for the base of the OT5 calculation. In addition to those with 5 valid domain scores, the base for OT5 includes children with less than five valid domain scores where the child has at least 1 valid domain score for which they are not developmentally on track. The PC's CtG target 4 reports the proportion of Aboriginal and Torres Strait Islander children assessed as developmentally on track in all five domains of the AEDC, where the total number of Aboriginal and Torres Strait Islander children in the first year of full-time school are used as the base of calculation (PC 2024).

Analysis cohort

From the full 2018 AEDC dataset we filtered to include children aged 4-6 at the time the instrument was conducted, with no identified special needs and at least one valid score in a domain. Children requiring special assistance because of chronic medical, physical or intellectually disabling conditions based on diagnoses (for example autism, cerebral palsy, Down syndrome) are defined in the AEDC as having 'special needs'. Teachers complete the AEDC for children with special needs but these children are not assigned a domain category (On track, At risk or Vulnerable) nor do they have a valid summary indicator score. Child age was determined by subtracting the age in months from the last month of AEDC data collection. These filters were applied using factors available in the AEDC dataset (Table 2).

Additionally, children who were not linked to MADIP and did not have a parent relationship recorded were filtered from the cohort.

This resulted in a final analytical cohort of 274,123 children (Table 2).

As a simplifying assumption, the only caring relationships analysed within the family unit were between the child and a linked parent(s). Informal or non-primary kinship care relationships were not analysed under the datasets and methodology. Parent relationship was sourced from the relationship scope dataset. Note that parent or carer highest educational attainment was taken directly from the AEDC rather than data associated with the linked parent, see Factors subsection.

This decision to focus on parent relationships was made for multiple reasons, including the difficulty reliably identifying primary caregivers, consistency with existing academic literature and our analyses frequently being centred on parent characteristics (for example household income, maternal age, parental education).

Based on the final analytical cohort of 274,123, the cohort number for individual analyses varies slightly due to the missing values in each domain of interest.

For the predictive modelling and G-computation and Inverse Propensity Weighting (IPW) modelling, multiple factors were analysed simultaneously (see *Predictive modelling* section and the Factors subsection within this *Methodology* section). Children with any factor with missing observations were removed. This resulted in a further reduction of the cohort (see Table 2). This does create a missing value bias that was a limitation of the study.

In particular, children on the AEDC who were unable to be linked to other datasets had higher rates of developmental vulnerability than the general population, and are expected to be less likely to use child care (because additional information collected in the CCMS data should increase the likelihood of linkage). This bias needs to be kept in mind when interpreting the unadjusted results, though is not expected to have the same impact on the modelling (predictive, G-computation or Inverse Propensity Weighting). See the Appendix for further details.

Table 2. Analytical cohort filters

Filter	Population size
Full AEDC cohort	308,873
Valid children (aged 4 to 6, no special needs)	294,605
MADIP-linked children	279,639
Children with parental data (analytical cohort)	274,123
Predictive modelling factors complete data	211,885
G-computation and IPW modelling attendance data	242,313
G-computation and IPW modelling quality data	232,693
G-computation and IPW modelling duration data	232,693

Source: Customised 'First Five Years' extract from the Multi-Agency Data Integration Project, 2021.

Notes: This table compares children from the 2018 cohort of the Australian Early Development Census.

Factors

Based on a literature review of factors known to impact child developmental vulnerability, we developed a list of factors to investigate. We derived these factors from the First Five Years MADIP Custom Extract. This process involved developing business rules to interpret the integrated administrative data and identifying data gaps and limitations of the proxy factors.

Cross-sectional effects

To understand the impact of certain characteristics across the first five years, many longitudinal factors were collapsed into cross-sectional factors.

Cross-sectional factors are time-insensitive – they aggregate all observations across the study period, presenting a summary value. Although longitudinal data was available within MADIP, given the inconsistency of collection across the different data sources and the extra technical complexity added, we aggregated these data to be represented by cross-sectional factors (for example, years in employment).

These cross-sectional factors included binary indicator factors as well as factors that were only measured at a single point in time (e.g. Census and AEDC data).

Child demographics

Gender

We identified children's gender using the AEDC gender factor. Only boys and girls were considered. Any other genders were excluded from analyses because of low sample size.

Language background other than English (LBOTE)

We identified children with LBOTE using the binary AEDC LBOTE factor.

Aboriginal and Torres Strait Islander status

We identified children as Aboriginal and Torres Strait Islander if they were recorded as Aboriginal and Torres Strait Islander in 50% or more of the data sources in which their Aboriginal and Torres Strait Islander status was known. There were four datasets used: AEDC, Census, DEX and CCMS. This method was developed in collaboration with the National Indigenous Australians Agency (NIAA) and follows the principles outlined in the National best practice guidelines (AIHW 2012).

Parent and family characteristics

Maternal age at birth

We identified maternal age at birth using the month and year of birth for both the mother and the child. Birth date information for mothers was sourced from the combined demographics dataset, while birth date information for children was sourced from the AEDC.

Country of parents' birth

We identified country of parents' birth using the country of birth indicator available in the combined demographics dataset. We categorised country of parents' birth into three groups: Australia, other Organisation for Economic Co-operation and Development (OECD), and non-OECD. When a child had parents from multiple categories, we prioritised categorisation in order of non-OECD, other OECD, Australia. This means that children with one parent born in an OECD country and one in Australia were categorised as OECD, while children with one parent born in a non-OECD country and one in an OECD country were classified as non-OECD.

Parent or carer highest educational attainment

We identified the highest level of educational attainment of a children's parents or carers using four factors from the AEDC: *Parent1School*, *Parent1PostSchool*, *Parent2School*, and *Parent2PostSchool*.

These factors recorded the level of high school and post-school education for up to two parents or carers for each child and were prefilled from school records collected when the child was enrolled into school. The order and other information prefilled into the AEDC did not give any identifying information including the gender of the parent or carer, whether they were a primary or secondary caregiver, or whether they were a parent or carer. While the AEDC provided the most current educational status of parents or carers available, it did not individually identify the parents or carers and therefore could not be directly linked to other sources of information about the parents.

For each child, we identified the highest level of education attained by each parent or carer. We then created two factors representing the highest educational attainment of the child's most educated parent or carer and the highest educational attainment of the child's least educated parent or carer.

The AEDC data contains a broad education level of Certificate level I to IV group, by treating four distinct education levels under the Australian Qualifications Framework (AQF) as one group (Department of Education, n.d.). The AQF does not include levels for Senior Secondary Certificate of Education qualifications; however, according to the International Standard Classification of Education 2011 to Australian Standard Classification of Education Concordance, Certificate I-II is below Year 12 (equivalent to Year 10) (DET 2019). The grouping of these levels with Certificate III-IV (higher than Year 12) may bias some results.

Unpaid child care

We identified a child as being exposed to unpaid child care if any parent self-reported providing unpaid child care for their own or other children in the past two weeks in the 2016 Census. Limitations of this definition are that unpaid child care may correspond to a different child and the Census only captures a single point in time.

Single parent household

We identified whether a child lived with one or two parents at the time closest to the completion of the AEDC. We considered the last known address in the period to be the effective address. Children with greater than two parents were out of scope and identified as unknown. To identify whether a child was living with one or two parents we made the following assumptions.

1. If a child and parent were living in the same Statistical Area Level 1 (SA1) they were assumed to be in the same household.
2. If a child only had one parent they were assumed to be part of that parent's household, regardless of whether they shared the same SA1.
3. If a child had exactly two parents and those parents lived in the same SA1 then the child was assumed to be part of that household.

We adjusted the single parent factor if tax data indicated presence of a spouse not otherwise identified. We then made a further adjustment to the factor based on DOMINO welfare parenting payments indicating a single or dual parent arrangement (Services Australia 2021).

Early childhood education and care

Formal child care attendance

We identified formal child care attendance using the Child Care Management System (CCMS) data. The CCMS reports quarterly charged hours of child care enrolment in long daycare centre (LDC), family day care (FDC) and outside school hours care (OSHC, noting this type of care was generally not included as the analysis focused on use before starting school). In our descriptive analysis of developmental vulnerability and rates of being developmentally on track by child care quality, duration or attendance, a child was identified as having attended formal child care if they had a record in the CCMS prior to the year they started school. A limitation of this definition is that it may include children who spent as little as one hour in child care and then never attended again.

In the modelling section, a simpler definition of child care was used whereby a child was identified as having attended child care with any record in the CCMS at all.

Note that the child care attendance here does not include ECEC attendance fully, with the lack of detailed information on government-run pre-school attendance.

Average weekly charged hours

We identified children's average weekly child care hours based on CCMS quarterly charged hours.

In other words, for each child, we found the average quarterly child care hours for all time, converted this value to average annual hours and divided through to find average daily hours. Finally, we multiplied by seven to find average weekly hours. Weekly hours are calculated by summing hours charged per quarter, dividing by the number of quarters the child had attendance data and multiplying by (4×7)/365.25.

The CCMS data included hours charged rather than hours attended. Business knowledge suggests that children attend for roughly 70% of the hours that families are charged. Therefore, it is likely that our *average weekly hours* factor is an overestimation of the amount of child care a child attended.

Quality

We identified child care quality using the Australian Children's Education & Care Quality Authority (ACECQA) National Quality Standards (NQS) (ACECQA 2021). The NQS ratings apply to all education and care services under the NQF and are not just a child care rating. The NQS data included quality ratings for seven quality areas (Education program and practice, Children's health and safety, Physical environment, Staffing arrangements, Relationship with children, Collaborative partnerships with families and communities, and Governance and leadership), in addition to an overall quality rating, for each child's most common care service provider per quarter. We created a quality factor that further simplified the overall quality rating into four categories:

1. Not yet at standard: quality ratings that did not meet the National Quality Standard. This included *Significant Improvement Required* and *Working Towards NQS ratings*.
2. At standard: quality ratings that met the National Quality Standard. This included the *Meeting NQS* rating.
3. Above standard: quality ratings that exceeded the National Quality Standard. This included *Exceeding NQS*, and *Excellent* ratings.

4. Provisional: not yet assessed under the NQF. At any one time a small proportion of services will only recently have been approved and may not have started operating or may have only been operating for a short period of time. In general, state and territory regulatory authorities will assess and rate newly approved services within 9-18 months of operations commencing.

From all the child care service providers with a non-provisional rating attended by a child, we identified the most frequently occurring quality rating for a child. Not all child care service providers had quality rating information. Our method for calculating modal quality is limited by CCMS and NQS data. NQS data for quality are recorded quarterly per child for only their most-attended child care provider. By linking the NQS data back to CCMS data we can also tell how many providers a child had per quarter and how many hours they attended in total. In the case that a child attended more than one provider, we cannot tell how the hours were divided between them.

Preschool attendance

Preschool school program is delivered to children in the year before they start full-time school. Different states and territories have different names for preschool services. In NSW, ACT and NT, the name preschool is used. In Victoria, Queensland, Tasmania and Western Australia, they are known as kindergartens, while in South Australia both preschool and kindergartens are used.

We defined a preschool attendance as having attended preschool as recorded by the teacher in the AEDC dataset or if the child had at least 600 hours of *Long Day Care* (LDC) in the CCMS in the year before school. Attendance for 600 hours at LDC in the year before school corresponds to the Universal Access National Partnership aim from both federal and state and territory governments to guarantee 600 hours of preschool or high-quality child care to all children (DESE 2021).

The CCMS tracks attendance hours on a quarterly basis. For the 2018 AEDC cohort, we summed the 2017 LDC quarterly hours in the CCMS. If a child had at least 600 hours of LDC in 2017 they were identified as having attended preschool.

The CCMS data included hours charged rather than hours attended, so this definition may overestimate the number of children who attend preschool through a centre-based day care provider.

Health and mental health

Child and parental mental ill-health

We defined child and parental mental ill-health using methodology advised by the Department of Health following the Australian Institute of Health and Welfare (AIHW) mental health report (AIHW 2021). This method used Medicare Benefits Schedule (MBS) and Pharmaceutical Benefits Scheme (PBS) data for items known to be associated with mental ill-health. An individual was identified as having mental ill-health if they accessed any mental health service or prescription in the relevant period (see below). If an individual had MBS or PBS data available but did not access specific items related to mental ill-health they were identified as not having mental ill-health. If an individual did not appear on MBS nor PBS they were identified as missing data.

The use of linked MBS and PBS data as a proxy for mental ill-health has limitations. A person accessing a mental health services or medication does not mean they have a diagnosed mental health condition. Not all individuals with mental health conditions will use government-supported services and no diagnostic information is captured in either dataset. Not all government-supported services are included, such as State and Territory or Primary Health Network services. While this is the best available measure we can derive from the available data, it is likely a measure of their access to services as well as whether people have mental ill-health. The relationship between having a mental ill-health and using MBS or PBS mental health services is likely to be different for different population groups: for example, people with lower incomes with high distress tend to be less likely to access MBS mental health services (Mental Health Australia 2023).

For MBS, we used Medicare items for:

- Psychiatric services;
- General Practitioner mental health services, such as a mental health plan;
- Clinical psychology services;
- Other psychology services; and
- Other mental health professional services.

For PBS, we used the Anatomical Therapeutic Classification codes:

- NO5A Antipsychotics;
- N05B Anxiolytics;
- NO5C Hypnotics and sedatives;
- N06A Antidepressants;
- N06B Psychostimulants and nootropics.

We identified a child as having mental ill-health if they accessed the above services between birth and 2018. We defined a child as having experienced parental mental ill-health if any of their parents had accessed the above services from one year before childbirth to 2018.

Years with parental mental ill-health

We identified long-term mental ill-health in parents by counting the number of years a parent had accessed mental health services as a proxy. If the parent accessed a mental health service in a given calendar year between one year before childbirth and 2018, this was identified as a year of access. The total years of access were then counted. For children with multiple parents, the parent with the longest mental ill-health period was taken as the relevant experience period for that child.

Age at mental ill-health onset

We defined a child's age-at-onset using their earliest recorded quarter of mental ill-health since birth. Parental mental ill-health age-at-onset was also calculated with respect to the child's age. Birth date information for children was sourced from the AEDC. For children with multiple parents, the parent with the earliest mental ill-health was taken as the age-at-onset.

Child and parent chronic ill-health

We identified child and parental chronic physical ill-health using a method developed by the Department of Health. This method used Medicare Benefits Schedule (MBS) data for items specific to chronic disease management. An individual was identified as having a chronic physical ill-health if they accessed any chronic disease management item (see below) during the relevant period. If an individual had MBS or PBS data available but did not access specific items related to chronic ill-health they were identified as not having ill-health. However, this factor should not be interpreted literally, as many (if not most) chronic health services would be delivered under other items, such as standard general practice consultations. If an individual did not appear on MBS nor PBS they were identified as missing data.

For MBS, we used Medicare items for:

• Chronic Disease Management;

• Allied Health.

Only MBS items for chronic ill-health related to physical ill-health were counted in the chronic health measure. There were a small number of cross-disciplinary MBS items that referred to both physical and mental ill-health and appeared in both derived factors. There are limitations with identifying chronic ill-health based only on use of MBS services for chronic disease management and allied health, as this approach may only capture a subset of those with chronic ill-health. Future work could explore using a rules-based approach for inferring conditions such as RX-Risk (Pratt et al. 2018), which uses PBS data.

We identified a child as having a chronic ill-health if they accessed the above services between birth and 2018. We defined a child as being exposed to parental chronic ill-health if any of their parents had accessed the above services from one year prior to childbirth to 2018.

Income and socio-economic status

Household income

Household or family income for each child was identified based on the sum of disposable parental income equivalised to the size of the household per financial year. Only parents identified in the relationship and location datasets who lived with the child that financial year were counted. We did not attempt to include income from siblings, other carers, or the children themselves in the household income.

The sum of the parents' income was aggregated from individual parent disposable incomes each financial year.

Disposable household income was used for consistency with external publications, including the ABS, Household, Income and Labour Dynamics in Australia (HILDA), OECD and the Smith family children's charity (ABS 2019; Wilkins et al. 2019; OECD 2021; Australian Council of Social Service 2018).

To calculate individual disposable income, we aggregated income from PIT, PAYG and DOMINO welfare for each individual parent closely following ABS methodology (ABS 2020; see *Appendix* for Individual income rules). Parents with no data across all three input sources were coded as missing.

For a minority of cases where we could only identify one parent's income and spousal income was present, we included the spouse income to the household total. For cases where two parents' income was reported we did not count spousal income, as we assumed the second income was the spouse. The Australian Tax Office (ATO) defines a spouse as a domestic partner residing together (ATO 2020) and based on this we included spouse dollars to the household total.

Spouse income data from PIT only had gross income factors available and did not include any net tax information so disposable income was not directly calculable. To estimate the tax paid, we took the average tax paid as a proportion of gross income for an individual in each tax bracket and applied that proportion to the spouse's gross income amount, to make an approximate net tax. We then subtracted the approximate net tax from the spouse's gross income to determine disposable spouse income.

Incomes are reported in June 2020 dollars, based on the Consumer Price Index (CPI) (ABS 2021c).

To identify parents that lived with the child we used shared SA1s for each financial year from the combined location dataset, as per the single-parent household derivation. The combined location dataset included time-sensitive, parent and child address data sourced from MEDB, DOMINO, PIT and Census 2016.

Household size was defined based on the count of parents (up to two parents) and children in the house each financial year. This number potentially changed each financial year.

We counted parents as per the relationship and shared SA1 data, as described above. When only one parent was identified we added a further parent if identified via spouse income. We made a further adjustment to the number of parents if DOMINO welfare parental payment data indicated this parent was in fact a single or couple parent (Services Australia 2021).

To count the number of children, we used the number of dependent children as per the PIT data each financial year. For multiple-parent families, the maximum value for the number of dependent children was taken as the number of children in the house. If this number was zero, we added one as the child in the AEDC cohort had to be counted.

To calculate household income, we summed the total disposable incomes of parents who lived with the child for that financial year, adding disposable spouse dollars for single parents. We then equivalised the total income by dividing it by the square root of household size (i.e. number of parents plus number of children).

The square root method for equivalising the household income was preferred as we did not have age data available for siblings, a requirement of other methods (OECD n.d.).

Household income decile

We identified the household income decile using household income across financial years from birth to completion of the AEDC. The household income decile was taken for each child for each financial year excluding the top 1 per cent and bottom 2 per cent of household incomes as per the top-tail ABS methodology (ABS 2019). To make a lifetime measure, we then took the most commonly occurring household income decile per child across all available financial years as our measure of household income decile.

Years with an employed parent

We identified children with employed parents for each financial year and created a lifetime factor that counted number of financial years a child had any employed parents. A child was deemed to have employed parents if any parent earned income by exertion (that is, excluding passive income streams) and lived with the child in a given financial year. Income by exertion was calculated using PIT and PAYG data following ABS methodology guidelines (ABS 2020; see *Appendix* for income rules). The total number of financial years the child had an employed parent was then counted between birth and the completion of the AEDC.

Neighbourhood SES

We identified the socio-economic status (SES) of the child's neighbourhood of residence using Socio-Economic Indexes for Areas (SEIFA) Index of Relative Advantage and Disadvantage (IRSAD). Developed by the ABS, IRSAD ranks geographical areas into 10 deciles based on relative social and economic advantage and disadvantage (ABS 2018a). We created an SES factor based on the Statistical Area Level 1 (SA1) of the child's residential address (ABS 2018b). When a child had no address but the parent(s) did, we used the SA1 of the parent's address. We defined neighbourhood SES as the child's lifetime minimum IRSAD decile, as it indicated at least some disadvantage during the child's life.

Welfare payments

We identified children who had parents who received welfare payments using the DOMINO dataset. As advised by Department of Social Services, welfare was categorised into non-exclusive payment groups based on receiving the following benefits:

- **Welfare receipt:** based on any benefit received.
- **Income Support:** Abstudy (Secondary/Tertiary), Age Pension, Austudy, Carer Payment, Disability Support Pension, Jobseeker Allowance, Newstart Allowance, Partner Allowance (historical), Parenting Payment Partnered, Parenting Payment Single, Partner Allowance, Sickness Allowance, Special Benefit, Sole Parent Pension, Widow Allowance, Wife Pension Age, Wife Pension DSP, Widow B Pension, Youth Allowance (apprentice), Youth Allowance (other), Youth Allowance (student), Youth Training Allowance.
- **Non-Income Support:** based on any benefit received that was not Income Support (above).
- **Rent Assistance:** Rent Assistance Family, Rent Assistance Parenting, Rent Assistance Newstart, Rent Assistance Pension, Rent Assistance Abstudy.
- **Disability Support Pension:** Disability Support Pension, Sickness Allowance.
- **Carer Payment:** Carer Payment, Carer Allowance.
- **Family Support:** Dad and Partner Pay, Double Orphan Pension, Assistance for Isolated Children, Family Tax Benefit Part A, Family Tax Benefit Part B, Parental Leave Pay, Parenting Payment Partnered, Parenting Payment Single, Sole Parent Pension.
- **Unemployment Payment:** Jobseeker Allowance, Newstart Allowance, Partner Allowance (historical), Partner Allowance, Special Benefit, Youth Allowance (other).
- **Student Benefits:** Abstudy (Schooling Applicant), Abstudy (Secondary/Tertiary), Austudy, Youth Allowance (apprentice), Youth Allowance (student), Youth Training Allowance.

- **Age Pension:** Age Pension, Widow Allowance, Wife Pension Age, Wife Pension DSP, Widow B Pension.

A child was identified as receiving a payment type if any parent received a relevant payment between the birth month of the child and the month of completion of the AEDC. Birth and AEDC completion date information was sourced from the AEDC.

Special child care benefit (SCCB)

We identified children whose parents received a SCCB using the following five SCCB identifiers in the CCMS, which indicated whether a child ever received any of the following:

- **Special Child Care Benefit:** A general factor for any child receiving special benefits. If a child had any of the following three benefits, then they also had this one too.
- **At-Risk Child Care Benefit:** A subset of the Special Child Care Benefit awarded to children who were identified as being at risk of negligence or abuse, as recognised by their child care provider.
- **Financial Hardship Child Care Benefit**: A subset of the Special Child Care Benefit awarded to children whose families were undergoing financial hardship, as recognised by their child care provider.
- **Grandparent Child Care Benefit:** A subset of the Special Child Care Benefit awarded to children who were cared for by their grandparents instead of their parents.
- **Jobs Education and Training (JET) Child Care fee assistance:** JET Child Care fee assistance provides extra help with child care costs for parents on income support while looking for work, studying or starting a job.

A note on exclusion of factors related to home learning

For our predictive and statistical modelling, and in any of our analysis, we opted to not include AEDC factors related to home learning environment because of potential post-treatment bias. While we understand there to be significant correlation between developmental vulnerability and other data items collected in the AEDC that do not directly contribute to the calculation of developmental vulnerability, we have chosen to exclude these factors from our analyses to avoid bias. We made this decision on advice from our academic partners at the Life Course Centre (personal communication, 5 July 2020):

> *The desirability of including these predictors is understandable. An extensive body of knowledge demonstrates the home learning environment, including activities such as reading to the child, and active engagement with schools is strongly associated with development of language, cognition, and social and emotional regulation. Thus, they are sources of intervention to improve child developmental outcomes and important confounders to consider when evaluating the effects of other covariates, such as childcare or socioeconomic status. The main reason we would advocate against using these measures, however, is that they are collected at the same time as the Australian Early Developmental Census (AEDC), whilst your covariates of interest are*

> *before the AEDC. Thus, they are post-treatment variables. Montgomery et al. (2018) provide detailed justifications for why this post-treatment adjustment should not be undertaken as it can bias estimates of causal (including descriptive or correlational) effects.*

The Vulnerable and Disadvantaged Children Project (Department of Education 2023), undertaken concurrently with this project, reported a strong positive correlation between developmental vulnerability and factors from Sections D and E of the AEDC, in particular, responses to the questions *E6 – has parent(s)/caregiver(s) who are actively engaged with the school in supporting their child's learning* and *E7 – is regularly read to/encouraged in his/her reading at home as far as you can tell*. Although the inclusion of these factors improved the performance of our models, we have excluded them for the reasons discussed above.

Analysis methodology

To investigate the factors influencing child development, we used descriptive analysis, predictive modelling, and statistical inference modelling. The differences between these techniques can be understood by the types of questions they are trying to answer.

Descriptive analysis seeks to understand how factors may be related to one another, for example *Are there differences in developmental vulnerability for children with different mental health status?*

Predictive modelling is used to achieve the best possible accuracy when trying to predict an outcome based on several factors, for example *How well can we predict developmental vulnerability?*

Statistical inference modelling answers 'what if?' questions, for example *What if we gave all children formal child care? How would this affect their development?*

Models used to assess the relative importance and relative effect sizes of our factors included logistic regression, random forest and causal inference. Our researchers used R (R Core Team 2020) to generate our results. RStudio (RStudio Team 2020) was used for R programming. Packages from these programs are also referenced throughout this document.

Descriptive analyses

Descriptive analysis was used to show trends between developmental vulnerability and other factors. This analysis compared frequencies and prevalence of developmental vulnerability between groups.

Predictive modelling using logistic regression and random forest

A stepwise logistic regression with Bayesian Information Criterion was used to select a reduced number of predictor variables. Using a both-direction regression, the algorithm combines both forward and backward steps, optimizing the model by adding significant variables and removing insignificant ones. This analysis used all available variables as listed in the methodology section, requiring all incomplete observations to be removed. As such, the sample size was reduced. This analysis was performed with R, using the *stats* package (R Core Team 2020).

The variables selected are used to generate two factors, namely relative importance of variables and the average marginal effects (AMEs), for their predictive power to model the developmental vulnerability of children.

A random forest model (Breiman 2001) was used to rank the importance of factors (selected from the stepwise logistic regress) for predicting developmental vulnerability for each gender (James et al. 2013). Permutation importance was calculated for each factor. To be *important* a factor must have a positive effect on the prediction accuracy. It involves calculating the drop in the predictive accuracy when the feature's values are replaced with randomly permuted values. This analysis was performed with R using the *caret* (Kuhn 2008) and *ranger* (Wright and Ziegler 2017) packages.

Average marginal effects (AMEs) for the indicators selected from the stepwise logistic regression average the effects of each indicator on developmental vulnerability across all children in the study. For example, the AME for 'preschool attendance' was calculated by taking the average of the effects of preschool across all children for all mental health, chronic health and other covariates to give an average percentage change in the likelihood of developmental vulnerability associated with this factor. This gave a broad overview of how each of the covariates impacted the likelihood of a child being developmentally vulnerable.

Estimate effect of child care using statistical inference techniques

The most robust approach to measure the effect of a treatment is to conduct a Randomised Controlled Trial (RCT), where people are randomly assigned to a treatment or control group and their outcomes measured afterwards. Due to the randomisation of treatment, any difference in outcome can be attributed to the different assignment rather than other differences in the populations (termed confounders), provided the sample size is large enough and the samples are truly random.

To establish a causal effect of child care, or any treatment, from observational data, additional analysis is needed which also requires the data to meet several identification criteria (Hernán and Robins 2020). Specifically, we need to meet:

- Exchangeability (no unmeasured confounding). This occurs when predictors of the outcome are equally distributed across treatment groups, as is assumed to occur in a randomised control trial. For observational studies this cannot be empirically tested, so researchers often try to approximate with enough confounders as informed by existing studies (Hernán 2012).
- Positivity: all individuals have a positive probability of receiving each treatment at every level of the confounders (Cole and Hernán 2008).
- Consistency: different ways of assigning a treatment have the same outcome (Cole and Frangakis 2009).
- No interference: a person receiving the treatment does not alter someone else's chance of receiving the treatment (Hernán 2012).

Once these criteria are met, an appropriate statistical analysis adjusts for confounding and is used to estimate the causal effect (Chatton et al. 2020). For example, regression modelling with inverse propensity weighting or G-computation.

However, statistical adjustment can only consider variables that are present in the data, unlike randomisation in an RCT, which should account for all other factors. Child characteristics, such as innate cognitive capabilities and parental characteristics such as quality of care provided, are difficult to measure or are not covered by available data. Thus, we would be unlikely to adjust for, or confirm the non-importance of, key confounding variables (violating exchangeability) and we cannot conclude causal effects for child care on developmental vulnerability.

Nonetheless, statistical methods are used to estimate the effect of child care on AEDC developmental vulnerabilities after adjusting for imbalance in confounding covariates. The results establish a better estimate of the effect of child care on developmental vulnerability; however, they require more research to be confirmed before the estimates could be considered causal.

We use two methods reducing confounding and compare their results: G-computation and Inverse Propensity Score Weighting (Tartaglia and Knapp (unpublished)). G-computation aims to compare treatments by estimating the outcome if every member of the study population were assigned to all the possible treatments available. This is achieved by first fitting a model to the data, then making copies of that data with each observation having a duplicate set of every different treatment. Predictions are then made as to the outcome for the copied data using the model fitted on the original data. Thus, for each person we get a result as though they were treated with every possible treatment. The difference between the treated/untreated outcome can then be used to estimate the average treatment effect and the risk ratio if everyone in the population is treated compared to a scenario where everyone is untreated. Inverse Propensity Weighting aims to use weights to adjust for the impact of covariates by upweighting or downweighting observations based on how common other factors are associated with the treatment group they are in. For example, if children from a high SES decile are more likely to attend child care they will be down-weighted proportionately, thereby adjusting for the effect of being in a high SES decile.

The two methods generally produce comparable results. In most cases in our report, only G-computation results were displayed. Studies on the effectiveness of inference techniques suggest that G-computation can be the best performing of the methods at discovering underlying causal relationships (Chatton et al. 2020). As such, G-computation results were given preference over Inverse Propensity Weighting results, noting the results were generally in agreement.

Selecting an appropriate set of confounders for which to control is critical for reliable effect estimation. This was based on available literature and in consultation with expert stakeholders. The covariates that were adjusted for in both causal inference techniques include whether a parent was a single parent or not, remoteness, child and parent mental health condition, child and parent chronic physical health condition, OECD country of origin of parents, Aboriginal or Torres Strait Islander status, highest education of parents, language background other than English, age of the mother at birth, household income, preschool attendance, years with an employed parent, socio-economic status of area of residence and gender.

References

ABS (Australian Bureau of Statistics) (2018a) *2033.0.55.001 – Census of Population and Housing: Socio-Economic Indexes for Areas (SEIFA), Australia, 2016*, ABS, accessed 5 January 2021.

ABS (2018b) *1270.0.55.005 – Australian Statistical Geography Standard (ASGS): Volume 5 - Remoteness Structure, July 2016*, ABS website, accessed 8 January 2021.

ABS (2019) *Household Income and Wealth, Australia methodology, 2017–18 financial year*, ABS website, accessed 11 February 2021.

ABS (2020) *Personal Income in Australia methodology, 2011–12 to 2017–18*, ABS website, accessed 23 February 2021.

ABS (2021a) *Multi–Agency Data Integration Project (MADIP)*, ABS website, accessed 21 December 2021.

ABS (2021b) *DataLab*, ABS website, accessed 21 December 2021.

ABS (2021c) *Consumer Price Index, Australia*, ABS website, accessed 23 February 2021.

ACECQA (Australian Children's Education and Care Quality Authority) (2021) *National Quality Standard*, ACECQA, accessed 10 December2021.

ACOSS (Australian Council of Social Service) (2018) *Inequality in Australia 2018 HTML*, ACOSS accessed 23 February 2021.

AEDC (Australian Early Development Census) (2019a) *About the AEDC domains*, AEDC website, accessed 18 February 2021.

AEDC (2019b*)* *Development of the AEDC*, AEDC website, accessed 23 February 2021.

AEDC (2019c) *FAQs for researchers*, AEDC website, accessed 23 February 2021.

AEDC (2022) *AEDC 2021 National Report,* AEDC website, accessed 3 September 2024.

AIHW (Australian Institute of Health and Welfare) (2012) *National best practice guidelines for data linkage activities relating to Aboriginal and Torres Strait Islander people*, catalogue number IHW 74, AIHW, Australian Government, accessed 3 September 2024.

AIHW (2021) *Mental health services in Australia*, AIHW website, accessed 23 February 2021.

ATO (Australian Taxation Office) (2020) *Spouse details – married or de facto 2020*, ATO website, accessed 23 February 2021.

Breiman L (2001) '*Random forests*', *Machine Learning*, 45(1):5–32, doi:10.1023/A:1010933404324.

Chatton A, Le Borgne F, Leyrat C, Gillaizeau F, Rousseau C, Barbin L, Laplaud D, Léger M, Giraudeau B and Foucher Y (2020) 'G–computation, propensity score–based methods, and targeted maximum likelihood estimator for causal inference with different covariates sets: a comparative simulation study', *Scientific Reports*, 10(1):9219, doi:10.1038/s41598–020–65917–x.

Cole SR and Frangakis CE (2009) 'The consistency statement in causal inference: a definition or an assumption?', *Epidemiology*, *20*(1), 3–5, doi:10.1097/EDE.0b013e31818ef366.

Cole SR, and Hernán MA (2008) 'Constructing inverse probability weights for marginal structural models' *American journal of epidemiology*, *168*(6), 656–664, doi:10.1093/aje/kwn164.

DET (Department of Education and Training) (2019) *International Standard Classification of Education 2011 (ISCED 2011) to Australian Standard Classification of Education (ASCED) Concordance*, accessed January 2024.

Department of Education (n.d.) *AQF Qualifications*, Australian Qualifications Framework website, accessed January 2024.

Department of Education (2023) Measuring vulnerability and disadvantage in early childhood data collections, Department of Education.

Hernán M (2012) 'Beyond exchangeability: the other conditions for causal inference in medical research', *Statistical Methods in Medical Research*, *21*(1), 3–5, doi:10.1177/0962280211398037.

Hernán M and Robins J (2020) *Causal inference: What if*. Chapman & Hall/CRC, Boca Raton.

James G, Witten D, Hastie T and Tibshirani R (2013) *An introduction to statistical learning,* Springer, New York.

Kuhn M (2008) *'Building predictive models in R using the caret package', Journal of Statistical Software*, 28(5):1-26, doi:10.18637/jss.v028.i05.

Mental Health Australia (2023) *Mapping mental health care*, Mental Health Australia, accessed 3 September 2024.

Montgomery JM, Nyhan B and Torres M (2018) 'How conditioning on post treatment variables can ruin your experiment and what to do about it', *American Journal of Political Science*, 62(3):760-775, doi:10.1111/ajps.12357.

OECD (Organisation for Economic Co-operation and Development) (n.d.) "Household income", in Society at a Glance 2024: OECD Social Indicators, OECD Publishing, doi:10.1787/c8568e7f-en.

OECD (2021) *Household disposable income (indicator)*, OECD website, accessed 23 February 2021, doi:10.1787/dd50eddd-en.

Pratt NL, Kerr M, Barratt JD, Kemp-Casey A, Ellett LMK, Ramsay E, Roughead EE (2018) 'The validity of the Rx-Risk Comorbidity Index using medicines mapped to the Anatomical Therapeutic Chemical (ATC) Classification System', BMJ Open 8:e021122, doi:10.1136/bmjopen-2017-021122.

Productivity Commission (PC) (2024) Socio-economic outcome area 4: Aboriginal and Torres Strait Islander children thrive in their early years, accessed 1 October 2024.

R Core Team (2020) *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria.

RStudio Team (2020) *RStudio: Integrated Development for R*, RStudio, PBC, Boston, MA.

Services Australia (2021) *Parenting Payment – who can get it*, Services Australia website, accessed 23 February 2021.

Tartaglia E and Knapp S (unpublished) 'Causal Inference in Education Data', *Australian Government Department of Education, Skills and Employment (DESE)*.

Wilkins R, Laß I, Butterworth P and Vera-Toscano E (2019) '[The Household, Income and Labour Dynamics in Australia Survey: Selected Findings from Waves 1 to 17]', *Melbourne Institute: Applied Economic & Social Research*, University of Melbourne, accessed 23 February 2021.

Wright MN and Ziegler A (2017) '[Ranger: A fast implementation of random forests for high dimensional data in C++ and R]', *Journal of Statistical Software*, 77(1):1–17, doi:10.18637/jss.v077.i01.

The First Five Years: What makes a difference?

5.2 Appendix

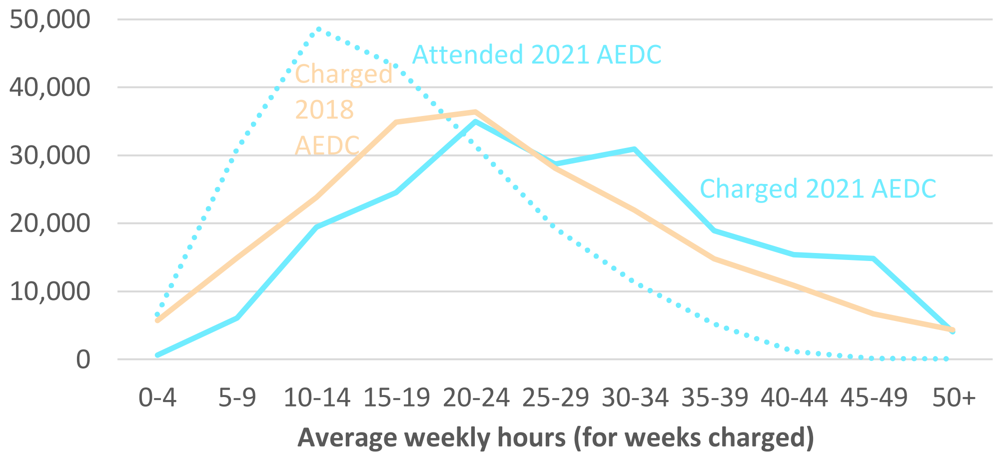
Appendix A: Note on child care charged and attended hours

It is well known that the hours children attend child care are generally less than charged hours. Child care centres typically charge for the whole time the centre is open to cover the cost of staffing and running the centre, while families rarely use the full session. Other reasons such as holidays and illness may also reduce the hours a child attends. Administrative data suggests that children attend for roughly 70 per cent of the hours that families are charged.

For the First Five Years project, the 2018 Australian Early Development Census (AEDC) is linked to the Child Care Management System (CCMS) data, where quarterly charged hours are recorded. Since July 2018, the new Child Care Subsidy (CCS) was introduced, and data on actual attendance at child care became available from 2019. We linked the CCS data with the 2021 AEDC to test the relationship between hours charged and hours attended, and the association with rates of children's developmental vulnerabilities.

The main goal of this subsection is to test the relationship between charged hours and attended hours, and then test whether the reported relationship between charged child care hours and children's development still holds for attended hours.

Figure 1: Number of people by average weekly charged hours and attended hours, 2018 and 2021 AEDC cohort.



Source: Custom Multi-Agency Data Integration Project (MADIP) extract Education, Skills and Employment National Data Asset (ESENDA), including AEDC data linked to CCS data, accessed 2023.

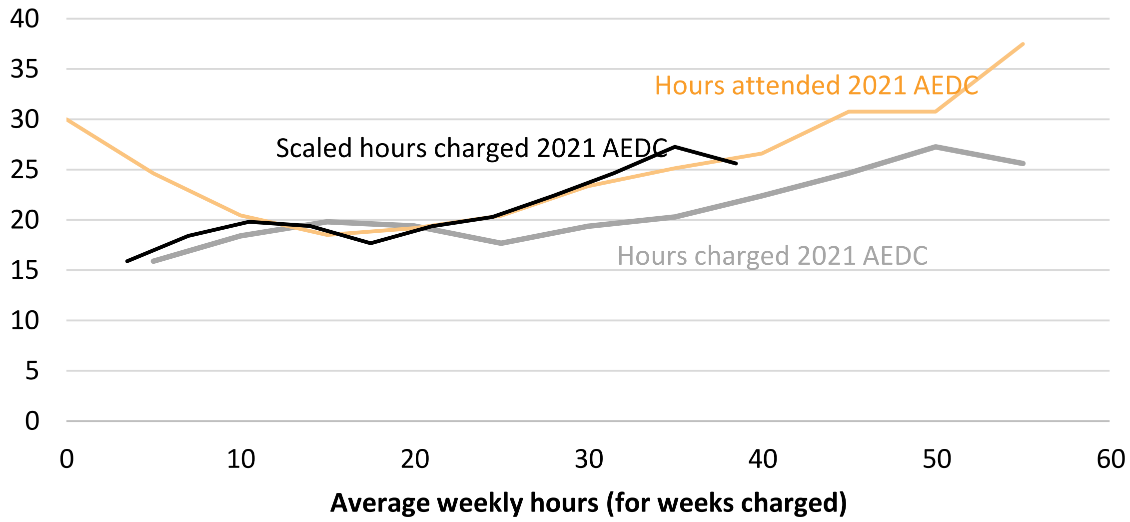
As shown in Figure 1, the distribution of hours families are charged is consistent between 2018 and 2021, with a slight shift towards higher amounts of charged hours in 2021. Most families are charged between 15 to 35 hours per week, equivalent to around 2 to 3 days per week. As expected, the actual hours attended are lower, with the peak at 10-14 hours of attendance per week.

Note that the data only capture attendance at a service eligible for the Child Care Subsidy, and as such does not include fully government-funded preschools.

CCS data contains both charged hours and attended hours, which allows the comparison of child developmental vulnerability for charged and attended hours. In Figure 2, the blue and grey lines show the rate of developmental vulnerability on one or more domains (DV1) based on charged hours and attended hours respectively. Low DV1 rates for the children with low charged hours are observed. However, the DV1 rates for the children attending short or no hours (while paying for them) are high. This low attendance could be due to family and/or health issues. A more detailed study in the future on the comparison of charged hours and attended hours will be useful to identify these extra factors.

We conducted a simple test of the validity of using hours charged as a proxy for hours attended, by comparing the distribution of rates of vulnerability for both measures. We scaled charged hours by 0.7 (in effect, assuming that hours attended represent 70 per cent of hours charged), and compared the resulting distribution of rates of developmental vulnerability to the true distribution based on actual hours attended (blue and black lines in Figure 2). Between 10 and 40 hours, the two curves are very similar, suggesting the charged hours is a reasonable proxy measure of the child care attendance (and developmental vulnerability for those attendance rates) for normal ranges of attendance. However, for the lowest and highest end of the attended hours, the results based on charged hours deviate and care should be taken when dealing with those cases.

Figure 2: Proportions (%) of children who were developmentally vulnerable on one or more domains (DV1), by average weekly charged hours and attended hours, 2021 AEDC cohort.



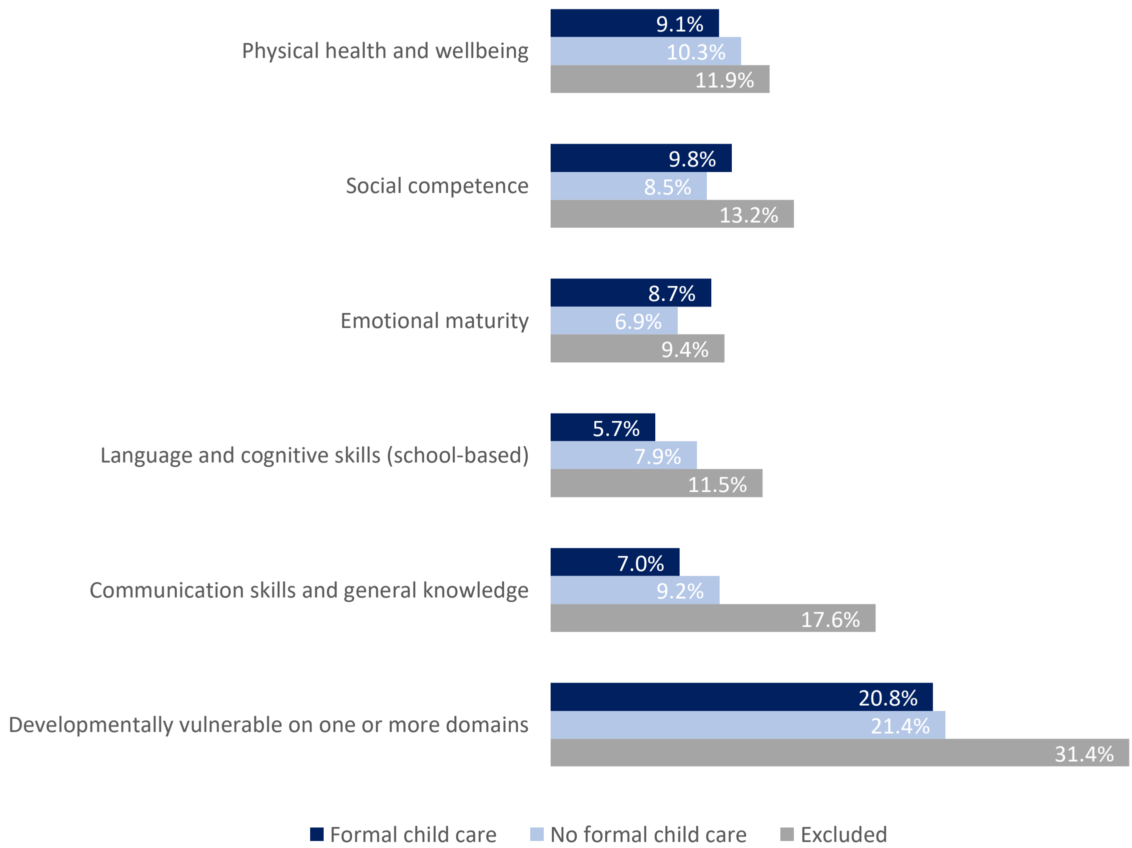
Source: Custom Multi-Agency Data Integration Project (MADIP) extract Education, Skills and Employment National Data Asset (ESENDA), including AEDC data linked to CCS data, accessed 2023.

Appendix B: Linkage bias

As shown in the *Methodology* paper (Table 2), the analytical cohort is smaller than the AEDC cohort due to linkage rates and exclusion of individuals without complete covariates.

For the descriptive analysis of formal child care attendance, we defined child care attendance as unknown if (1) the child could not be linked to the MADIP data, and (2) if the child did not have information of a primary carer. As can be seen from Figure 3, this group of children were more likely to be developmentally vulnerable in all domains compared to the analytical cohort, including both those who did and did not attend child care. Thus, the analytical cohort is biased towards those who are less likely to be developmentally vulnerable.

Figure 3. Proportion (%) of children who are developmentally vulnerable on different domains, for the analytical cohort attending child care and not attending child care and those excluded from the analytical cohort, 2018 AEDC cohort.



Source: Custom Multi-Agency Data Integration Project (MADIP) extract Education, Skills and Employment National Data Asset (ESENDA), including AEDC data linked to CCS data, accessed 2023.

Notes: Analytical cohort with valid DV1 score (including child care and no child care): N=272,616; Rest of cohort N=20,286

For the predictive modelling and statistical inference work, any child with one or more missing values was removed from the analytical cohort, resulting in a further reduced cohort of 211,885. A bias can be estimated for different variables as the difference between the data with no observations removed and the data with all observations with missing data removed. The study of the bias shows that the individuals with missing data were more likely to be developmentally vulnerable and disadvantaged in other ways, including being more likely to live in a low socio-economic area, less likely to have employed parents or parents with a bachelor degree. The results of our predictive modelling and inference work are based on a cohort that is biased towards being representative of less disadvantaged individuals.

Appendix C: Detailed statistical inference results for chapter 3.4

Detailed analysis results such as confidence intervals are available upon request.

Effect of child care duration analysis:

Sample sizes for statistical inference analysis on duration in child care							
	Average weekly hours						
Total hours	0 (control)	<10	10-19	20-29	30-39	40-49	50
0 (control)	55661						
1-2000		13456	25517	5933	877	101	
2001-4000		56	19839	21416	3124	571	1569
4001-6000			2012	25841	7631	1364	2137
6001-8000				6447	15736	2845	80
8001-10000				56	7117	6284	168
10001-12000					397	4780	269
12000						846	563

Appendix D: Individual income rules

We calculated individual incomes each financial year closely following an ABS methodology (ABS 2020; see Methodology for Household income rules). Income data sourced from Personal Income Tax (PIT), Pay As You Go summaries (PAYG) and Data Over Multiple Individual Occurrences (DOMINO).

Employee/Exertion:

- Salary or wages (PIT) +
- Allowances earnings tips directors’ fees etc (PIT) +
- Employment termination payments taxable component (PIT) +
- Attributed personal services income (PIT) +
- Total reportable fringe benefits amount (PIT) +
- Reportable employer superannuation contributions (PIT) +
- Other net foreign employment income (PIT) +
- Foreign source income exempt foreign employment income (PIT) +
- Lump sum A (PAYG) +
- Lump sum B (PAYG) +
- Lump sum D (PAYG) +
- Lump sum E (PAYG)

When PIT data is unavailable or zero, use PAYG only definition:

- Lump sum A (PAYG) +
- Lump sum B (PAYG) +
- Lump sum D (PAYG) +
- Lump sum E (PAYG) +
- Total allowances (PAYG) +
- Gross Payment Amount (PAYG) +
- Exempt foreign employment income (PAYG)

Business:

- Net income or loss from business - primary production (PIT) +
- Net income or loss from business - non-primary production (PIT) +
- Distribution from trusts primary production (PIT) +
- Net PSI (PIT) +
- Distribution from partnerships less foreign income non primary production (PIT) +
- Distribution from partnerships primary production (PIT)

Investment:

- Gross interest (PIT) +
- Dividends unfranked (PIT) +
- Dividends franked (PIT) +
- Dividends franking credit (PIT) +
- Share of net income from trusts less net capital gains and foreign income non primary production (PIT) +
- Franked distributions from trusts non primary production (PIT) +
- Foreign source income Australian franking credits from a NZ company (PIT) +
- Foreign source income net foreign rent (PIT) +
- Rent net rent (PIT)

Superannuation:

- Australian annuities and superannuation income streams taxable component taxed element (PIT) +
- Australian annuities and superannuation income streams taxable component untaxed element (PIT) +
- Australian annuities and superannuation income streams lump sum in arrears taxable component taxed element (PIT) +

- Australian annuities and superannuation income streams lump sum in arrears taxable component untaxed element (PIT) +
- Australian superannuation lump sum payments taxed element (PIT) +
- Australian superannuation lump sum payments untaxed element (PIT) +
- Bonuses from life insurance companies and friendly societies (PIT)

Other:

- Foreign source income net foreign pension or annuity without UPP (PIT) +
- Foreign source income net foreign pension or annuity with UPP (PIT) +
- Foreign entities controlled foreign company income (PIT) +
- Foreign entities Transferor trust income (PIT) +
- Foreign source income other net foreign source income (PIT) +
- Other income category 1 (PIT) +
- Other income category 2 (PIT)

Welfare:

- Australian Government pensions and allowances (PIT)

When PIT data is unavailable or zero, use DOMINO based derivation for total welfare payments:

- Component amount paid (daily), multiplied by days received per financial year. All payment types summed. (DOMINO)

Tax:

- Net tax (PIT) +
- Medicare levy (PIT) +
- Medicare levy surcharge (PIT) -
- Medical expenses offset available

When PIT data is unavailable or zero, use PAYG only definition:

- Tax Withheld Amount (PAYG)

Gross:

- Employee/Exertion (derived) +
- Business (derived) +
- Investment (derived) +
- Superannuation (derived) +
- Welfare (derived)

When PIT data is unavailable or zero, use PAYG and DOMINO based definitions:

- Employee/Exertion (derived) +
- Reportable fringe benefits amount (PAYG) +
- Tax-free component (PAYG) +
- Taxable component (PAYG) +
- Welfare (derived)

Disposable:

- Gross (derived) -
- Total deductions (PIT) -
- Tax (derived)

When PIT data is unavailable or zero, use PAYG and DOMINO based definitions:

- Gross (derived) -
- Tax (derived)

Spousal gross:

- Your spouse's taxable income (PIT) +
- Amount of Australian Government pensions and allowances that your spouse received (PIT) +
- Amount of exempt pension income your spouse received (PIT) +
- Amount of your spouse's reportable superannuation contributions (PIT) +
- Your spouse's target foreign income (PIT) +
- Your spouse's total net investment loss (PIT) +
- Your spouse's total reportable fringe benefits amounts

References

ABS (Australian Bureau of Statistics) (2020) *Personal Income in Australia methodology, 2011—12 to 2017—18*, ABS website, accessed 23 February 2021.