Australian Government
Department of Education

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).

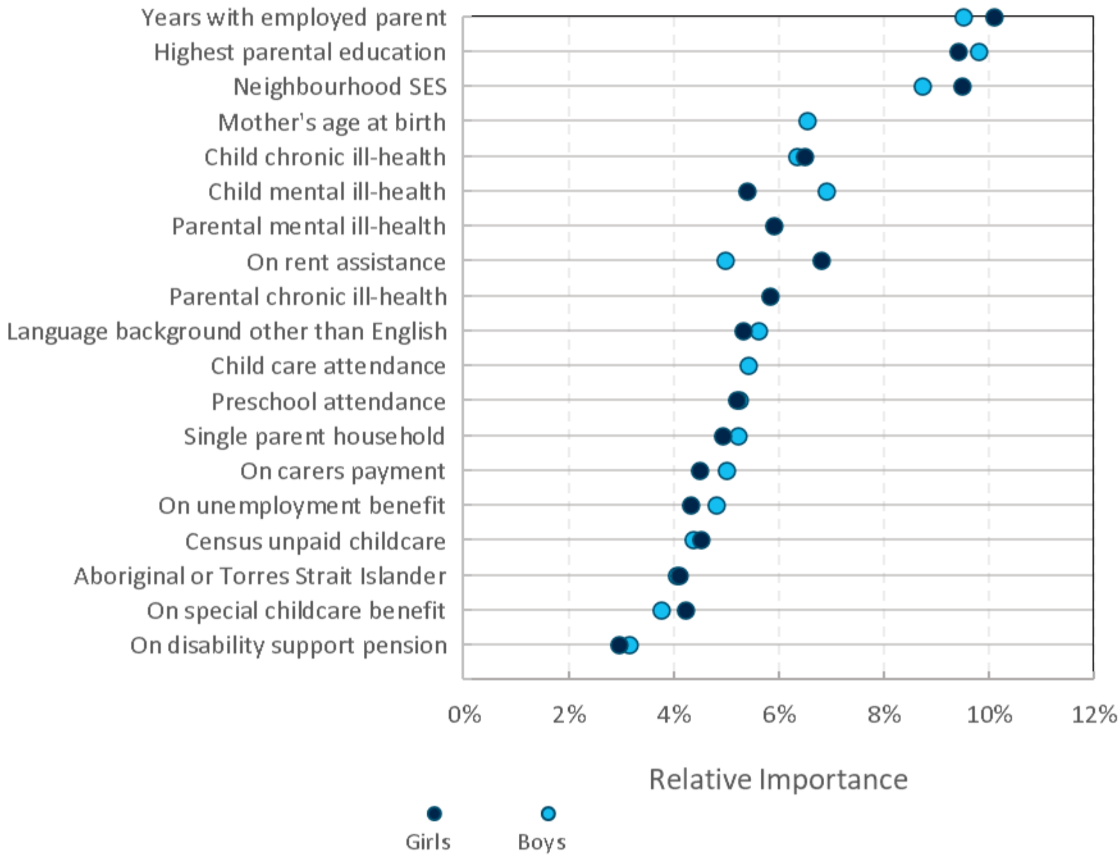
Which factors are important in predicting developmental vulnerability?

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?

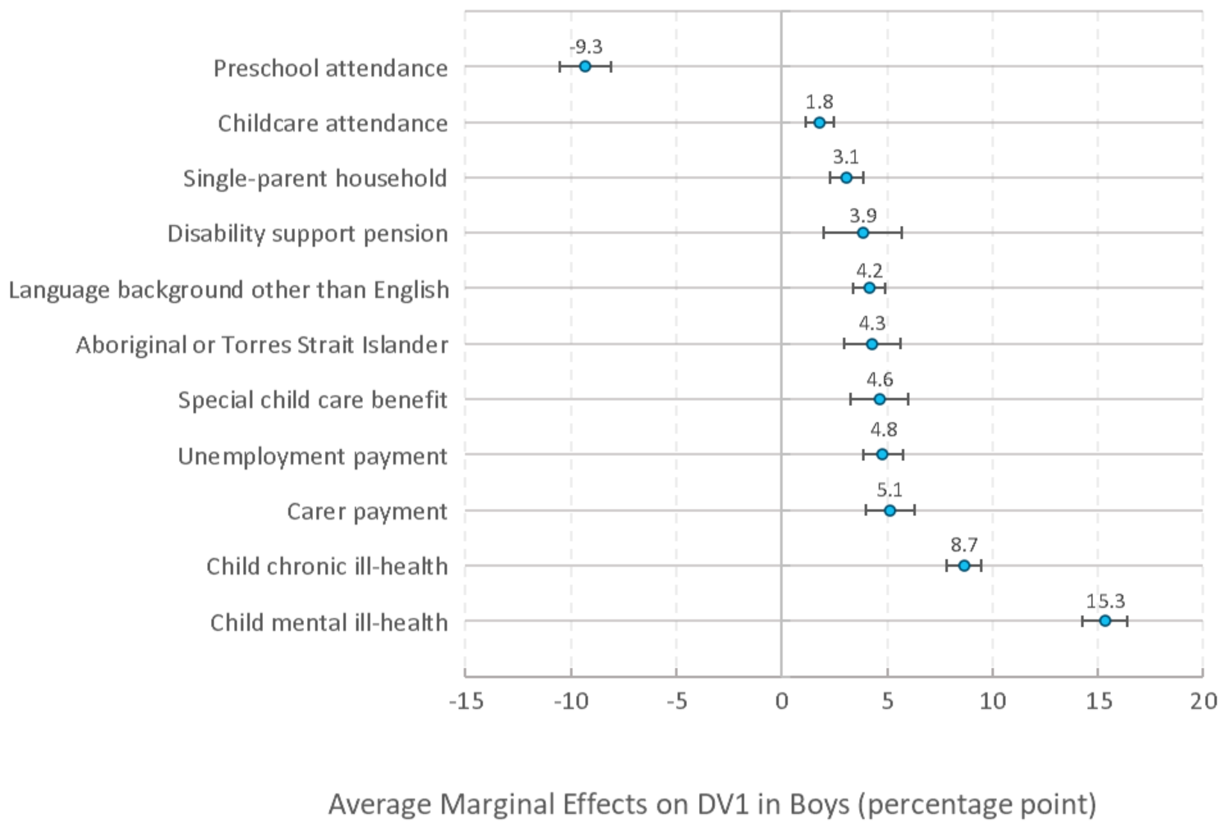
We calculated average marginal effects (AMEs) for the selected factors from the stepwise logistic regression. The AME gives an indication of a factor's impact on the likelihood of a child being developmentally vulnerable, compared to the reference group. For example, for the preschool attendance variable, 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.



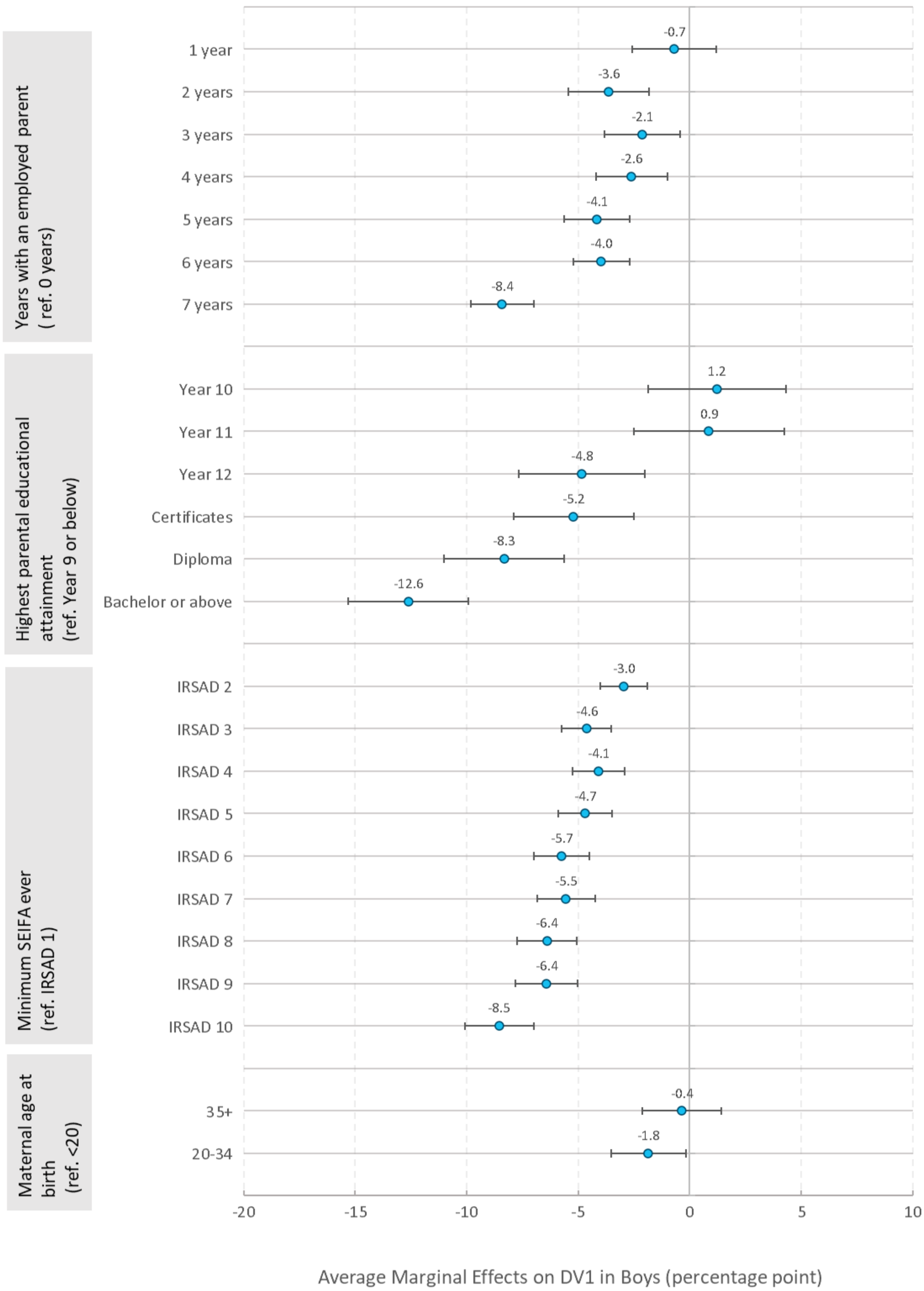
Source: Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.

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.

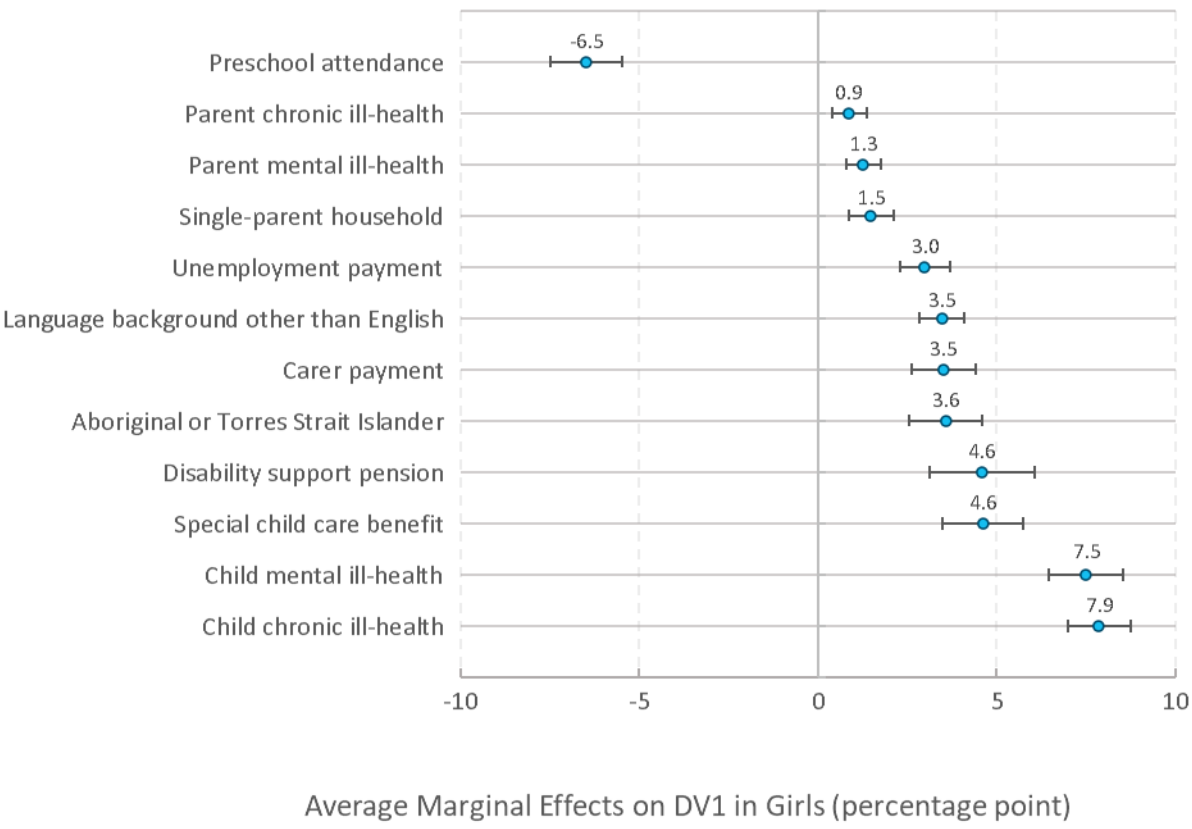


Source: Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.

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.

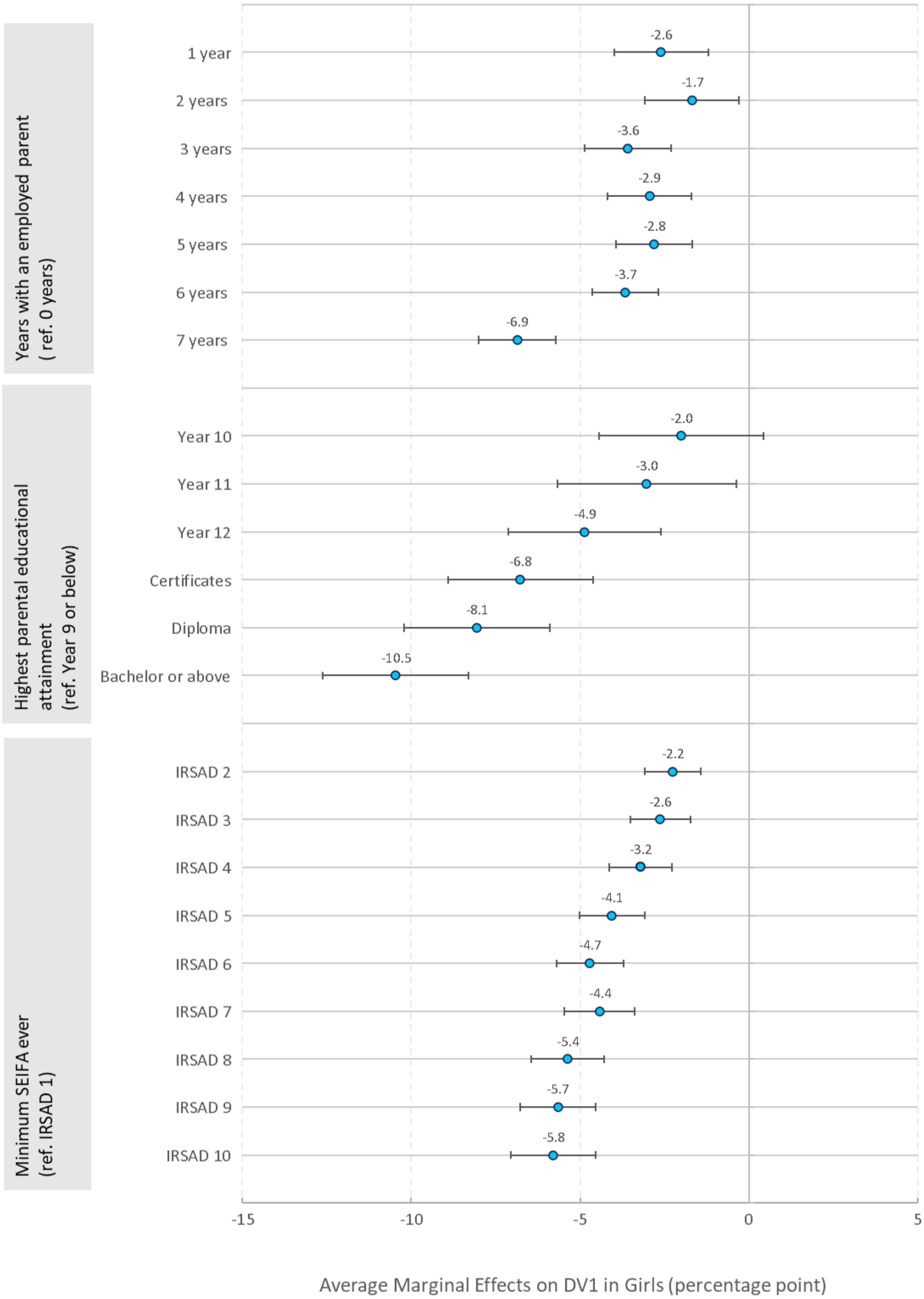


Source: Customised 'First Five Years' extract from the Multi-Agency Data Integration Project.

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.