EVIDENCE CENTRE
TE POKAPŪ TAUNAKITANGA

Evaluation of the Family Start programme

Report on findings of the impact evaluation

April 2021

ALLEN+CLARKE

ORANGA TAMARIKI
Ministry for Children

EVIDENCE CENTRE
TE POKAPŪ TAUNAKITANGA

Aurora Centre, 56 The Terrace, Wellington

The Oranga Tamariki Evidence Centre works to build the evidence base that helps us better understand wellbeing and what works to improve outcomes for New Zealand's children, young people and their whānau.

Email: research@ot.govt.nz

Authors: Nick Preval (Oranga Tamariki Evidence Centre/Allen + Clarke), Eyal Apatov (Oranga Tamariki Evidence Centre), and Jacinta Cording (University of Canterbury/Allen + Clarke).

Published: April 2021

ISBN: 978-0-9951443-9-2

If you need this material in a different version, please email us at research@ot.govt.nz and we will provide it for you.

Citation guidance:

This report can be referenced as: Preval, N., Apatov, E., & Cording, J. (2021). *Evaluation of the Family Start programme: Report on findings of the impact evaluation.* Wellington, New Zealand: Oranga Tamariki—Ministry for Children.

Oranga Tamariki disclaimer:

Oranga Tamariki has made every effort to ensure the information in this report is reliable but does not guarantee its accuracy and does not accept liability for any errors.

Statistics New Zealand IDI disclaimer:

The results in this report are not official statistics, they have been created for research purposes from the Integrated Data Infrastructure (IDI) managed by Statistics New Zealand.

The opinions, findings, recommendations and conclusions expressed in this report are those of the authors, not Statistics NZ.

ORANGA TAMARIKI
Ministry for Children

Access to the anonymised data used in this study was provided by Statistics NZ in accordance with security and confidentiality provisions of the Statistics Act 1975. Only people authorised by the Statistics Act 1975 are allowed to see data about a particular person, household, business or organisation and the results in these tables have been confidentialised to protect these groups from identification.

Careful consideration has been given to the privacy, security and confidentiality issues associated with using administrative and survey data in the IDI. Further detail can be found in the privacy impact assessment for the Integrated Data Infrastructure available from www.stats.govt.nz.

Acknowledgements:

Special thanks to Sarah Crichton, Principal Advisor, The Treasury, for her invaluable and extensive input into the design and implementation of the analytical approach.

He Awa Whiria advisory group:

Kaitiaki Māori

Andrew Sporle	University of Auckland

Kaitiaki Pasifika

Dr Jesse Kokaua	Centre for Pacific Health, Va'a O Tautai, University of Otago

Facilitator

Shirley Simmonds (Raukawa, Ngā Pūhi) Kaupapa Māori Researcher

Technical advisor

Dr Dean Hyslop	Senior Fellow, Motu Economic and Public Policy Research

Project team

Marnie Carter	Allen + Clarke
Dr Jessie Wilson	Allen + Clarke
Dr Rick Audas	University of Otago

Evaluation Advisory Group (EAG):

Davina Jones	EAG Chair, Manager–Evidence Centre, Oranga Tamariki
Moira Wilson	Senior Analyst, Ministry of Social Development
Sarah Crichton	Principal Advisor, The Treasury
Valmai Copeland	Principal Advisor, Oranga Tamariki
Janet Dean	Senior Advisor, Pa Harakeke Family, Oranga Tamariki
Jean Mitaera	Programme Leader – Bachelor of Social Work, Whitireia New Zealand
Judy Matai'a	Chief Executive, Anglican Trust for Women and Children (ATWC)
Nicole Coupe	Chief Executive, Kirikiriroa Family Services Trust
Gail Gillon	Director, Child Well-Being Research Institute, University of Canterbury
Carra Hammon	Manager, Investment Planning and Support, Te Puni Kokiri

Authors of the 2016 and 2017 evaluations of Family Start who shared code and provided valuable feedback:

Professor Rhema Vaithianathan	School of Economics, Auckland University of Technology
Moira Wilson	Senior Analyst, Ministry of Social Development

Statistics New Zealand IDI Customer Liaison Team:

Thanks to all of the members of the team who we interacted with in any capacity, for their professional, timely and friendly support.

CONTENTS

LIST OF TABLES

EXECUTIVE SUMMARY

This report provides findings from an impact evaluation of the Family Start programme ('Family Start'), a voluntary home-visiting programme that supports whānau/families (henceforth referred to as 'whānau') that struggle with health and social challenges to improve children's health, learning and relationships, whānau circumstances, environment and safety.

The findings of this report complement those from the Family Start process evaluation,[1] and inform the Family Start evaluation synthesis report.[2]

Evaluation purpose and focus

The purpose of this impact evaluation was to assess the impact that Family Start participation is having on the wellbeing of New Zealand children and their whānau. The evaluation examined a range of health, educational, and social outcomes for children, including separate analyses for Māori and Pasifika children.

The evaluation was informed by a Bridging Cultural Perspectives approach which comprises the He Awa Whiria and Negotiated Spaces models.[3] Reviewers representing the three knowledge streams (Māori, Pasifika, Pākehā) worked together to assist with the interpretation of the findings at an aggregate level, and through the lens of each worldview.

Evaluation methods

The impact evaluation assessed the impact of Family Start on outcomes for children and their whānau across three broad outcome domains: post-neonatal mortality; health and education-related outcomes; and child protection outcomes.

The evaluation was conducted using data from Statistics New Zealand's Integrated Data Infrastructure (IDI), utilising two quasi-experimental methods: Propensity Score Matching (PSM) and Difference-in-Differences (DiD). These methods provide complementary approaches to assessing impact, with different strengths and weaknesses in the context of this programme.

PSM provides an individual-level analysis of the impacts of Family Start by comparing the outcomes of participants with those of matched non-participants who had similar observable characteristics. For PSM to be valid (and provide unbiased estimates), all characteristics that predict participation and affect outcomes of interest need to be included in the model. This is unlikely to be the case in the current study due to the variety of referral channels, a lack of

[1] Oranga Tamariki & Allen + Clarke (2020). Evaluation of the Family Start Programme: Report on findings of the process evaluation. Oranga Tamariki—Ministry for Children. https://orangatamariki.govt.nz/assets/Uploads/About-us/Research/Latest-research/Family-Start/Evaluation-of-the-Family-Start-programme.pdf

[2] Carter, M., & Cording, J. (2021). *Evaluation of the Family Start Programme: Synthesis of process and impact evaluation findings*. Oranga Tamariki—Ministry for Children.

[3] Superu (2018). Bridging Cultural Perspectives. http://www.superu.govt.nz/sites/default/files/Bridging%20Cultural%20Perspectives%20FINAL.pdf

administrative data on many of the referral criteria, and because participation in Family Start is voluntary. The resultant bias could be in either direction, with the extent of bias unknown.[4]

DiD provides an area-level analysis and compares changes in average outcomes in TLAs before and after the programme became available, relative to the changes experienced in TLAs where Family Start was not available over the same time period. While DiD is generally considered the stronger of the two methods (because it controls for unobserved factors that remain constant over time), we faced several challenges when applying this method in this study. Most significantly, we were not able to accurately identify the Family Start target group in the data (i.e. the children who are more likely to participate in the programme), with only 18% of children in the target group identified using our criteria actually participating in Family Start[5]. Additionally, some outcomes were rarely observed (e.g., infant mortality) and the observation period prior to Family Start being expanded in 2005-2007 was relatively short, hence the number of outcome occurrences were low. These factors all materially decreased the likelihood of detecting the effects of Family Start using DiD. There was also evidence that time trends in outcomes were not similar across TLAs before Family Start was introduced, meaning that estimated impacts from the DiD analysis could not be attributed to the programme.

The PSM analysis examined outcomes across participants from TLAs that enrolled in the programme between 2009 and 2015. By contrast, the DiD analysis examined the 2003 to 2015 period, focusing on outcomes in the 14 TLAs where the programme became available between 2005 and 2007. Both methods examined outcomes within the child's first, second, and sixth years.

Key findings

Post-neonatal mortality outcomes

PSM analyses estimated significant reductions in mortality in the first year of life (but not in the second or sixth years). In the first year of life, Family Start participation is estimated to have reduced overall **post-neonatal mortality** among participants by 42%, or 1.2 deaths per 1,000 children, reduced **SUDI-related deaths** among participants by 51%, or 0.7 deaths per 1,000 children, and reduced **injury-related deaths** among participants by 67%, or 0.6 deaths per 1,000 children.

Significant reductions in **post-neonatal mortality due to injury**, and **post-neonatal SUDI** were estimated for Māori participants of 0.8 and 1.0 deaths per 1,000 children, which equate to reductions of 59% and 63%, respectively.[6] For Pasifika participants, a significant reduction was estimated in overall **post-neonatal mortality** of 2.0 deaths per 1,000 children, or a reduction of 62%.

While estimates from DiD analyses consistently demonstrated a reduction in the mortality measures examined, the estimates were not statistically significant. The results did not show that the introduction of Family Start had a statistically significant impact on the risk of **post-neonatal**

[4] For example, if participants are more motivated than non-participants, and outcomes are on average better for more motivated whānau, then the PSM estimates will capture both the benefits from the programme, and the benefits from the additional motivation amongst participants. In that instance, the benefit from the programme will be over-stated. By contrast, many whānau are referred to the programme from social services, indicating that they may face additional challenges, which could lead to under-stating any benefits from participation.

[5] PSM provides an estimate of the direct impact on participants, whereas DiD provides an estimate of the impact on those in the target group (which includes both participants and non-participants). If the impact of Family Start is only experienced by participants, DiD estimates are expected to be around a fifth of those obtained in PSM.

[6] The reduction in overall post-neonatal mortality was significant at the 10% level.

mortality for children in the target group (i.e. children most likely to participate in Family Start).[7] As discussed above, the proportion of children in the target group used for the DiD analysis who actually received Family Start was less than one fifth, making it challenging to detect statistically significant effects. The confidence intervals around the DiD estimates were wide, indicating that the size of the estimated impacts is uncertain, reflecting the rarity of **post-neonatal mortality** and the associated sample size issues.

Health and education-related outcomes

PSM estimates indicated that Family Start significantly increased participants' likelihood of being **enrolled with a Primary Health Organisation (PHO)**, being **fully immunised at each milestone age**, and **attending a Before School Check (B4SC)**. These outcomes were also found for Māori and Pasifika children, and suggest that Family Start is making progress towards meeting short-term outcomes related to the 'Child's health and safety' domain outlined in its Theory of Change model, specifically that the child and their whānau are enrolled with PHO, and that the child's immunisations are up to date.

Increasing engagement with ECE is also one of Family Start's goals. However, there was no statistically significant impact on the likelihood of having ever **enrolled with an Early Childhood Education (ECE) provider** at any time by age six. Measures focused on the age of first enrolment or duration and intensity of ECE enrolment were not examined. Two additional findings were more ambiguous: PSM analyses limited to B4SC attendees found that Family Start participants who attended the B4SC were more likely to have **significant and non-significant issues identified in the B4SC**, and all participants' mothers were more likely to receive publicly funded **mental health and addiction services**.

Although some of the health-related impacts could be interpreted as negative (in that they may indicate deteriorating child/whānau circumstances), they could equally reflect the possibility that Family Start increased the rate of identification and provision of support for the physical and mental health issues of children and mothers. This interpretation is in-line with the programme's short-term goals of identifying (and addressing) safety and health issues of children and identifying and providing access to services for mothers with post-natal depression. For the B4SC it is possible that the success of Family Start in increasing enrolment has resulted in the participation of families that face greater challenges and thus we observe a greater proportion of children with significant and non-significant issues. This could lead to biased estimates since PSM matches children's characteristics at (and before) the time of birth, and not at the time the assessment was made.

The DiD analysis was limited to the impacts on PHO enrolment and maternal mental health service use due to data availability issues. The analysis did not find any differences in PHO enrolment rates. For mental health service use, the analysis suggested lower service use amongst mothers during the child's first two years of life in TLAs where Family Start was available. However, these findings could not be attributed to the programme as they appeared to be a continuation of pre-existing trends.

[7] Of note, reductions in overall post-neonatal mortality and post-neonatal mortality due to injury were statistically significant when changing the start of the study period from 2003q1 to 2004q3 (as in Vaithianathan et al., 2016).

Child protection outcomes

Overall, the PSM results suggested that Family Start participants were more likely to **interact with Oranga Tamariki** (e.g., Reports of Concern, Care Placements, reports of Family Violence events). Participants were also more likely to be **hospitalised for maltreatment related injuries and long-bone fractures**. Notably, the likelihood of these events occurring was greater during the child's first two years, compared to the sixth.

Again, while these may be perceived as concerning findings (as they indicate deterioration in child/whānau circumstances), they may reflect the programme's success in identifying and addressing family violence, alcohol and drug misuse, and child health and safety issues. This 'safeguarding' effect was discussed in Vaithianathan et al. (2016), where greater interaction with Oranga Tamariki results from the obligation of Family Start workers to report any child protection-related issues to Oranga Tamariki.[8] Similarly, hospitalisations outcomes may reflect an increased likelihood of children receiving required treatment, rather than an increase in instances of maltreatment. As discussed by Vaithianathan et al. (2016), these results could also capture children referred to Family Start following interaction with child protection and/or health services (i.e. reverse causality).[9] Therefore, we caution against interpreting the child protection-related findings as a negative effect of participation in Family Start.

The DiD analysis found one statistically significant impact for child protection outcomes: a greater likelihood of children from TLAs with Family Start (or their siblings) being recorded in a **Family Violence notification by Police to Oranga Tamariki** in the first year of life. Again, this result is not surprising since a goal of the programme is to address family violence and alcohol and drug misuse. For children in their sixth year of life, the DiD analysis suggests a reduction in such interactions, although the reductions observed were either not statistically significant, or are a continuation of pre-existing trends and thus are not attributable to the programme.

Replication of Vaithianathan et al. (2016)

This evaluation aimed to enhance and strengthen the existing body of evidence relating to the effectiveness of Family Start. The evaluation built on an earlier impact evaluation (Vaithianathan et al., 2016[10]) by using very similar outcome measures and population groups, but introducing modifications to the original modelling approaches. Our analysis also benefited from access to data over a longer timeframe.[11]

---

[8] Throughout this document, Oranga Tamariki is used to refer to the Care and Protection agency in place at the time. Prior to 2017, Child Youth and Family (CYF) was responsible for the care and protection of children.

[9] The extent to which this affects the results could be tested by matching participants to non-participants at the time of enrolment (rather than birth). However, this approach will not separate the effects of the programme from that of 'safeguarding'. 'Safeguarding' effects may also explain the greater likelihood of participants recording these outcomes at a later age, if recording these outcomes at an earlier age means that these children will be more closely monitored. This could be partially tested by re-estimating year 6 child protection outcome via PSM, and including these same outcomes in year 1 and 2 in the list of (exact) matching variables.

[10] Vaithianathan, R., Wilson, M., Maloney, T., & Baird, S. (2016). *The impact of the Family Start home visiting programme on outcomes for mothers and children. A quasi-experimental study*. Ministry of Social Development. https://www.msd.govt.nz/about-msd-and-our-work/publications-resources/evaluation/family-start-outcomes-study/index.html

[11] For example, PSM analyses included children born between 2009 and 2015 (compared to 2009-2011 in the previous evaluation). DiD analyses included children born between 2003 and 2015 (compared to 2004q3-2011q4 in the previous evaluation).

One of the goals of this analysis was to explore whether it was possible to replicate and hence validate the results of Vaithianathan et al.'s original analyses when using a comparable study period, data, and methods. We were able to replicate the PSM results very closely, including confirming statistically significant reductions in overall post-neonatal mortality and deaths from SUDI (Sudden Unexpected Death in Infancy) and injury in the first year of life. We were not able to replicate the key DiD finding of a statistically significant reduction in post-neonatal mortality in the first year of life when using our best attempt to replicate their model. However, we estimated (somewhat smaller) statistically significant reductions in post-neonatal mortality when applying various modifications to the model. In addition, some of the modified specifications also estimated a statistically significant reduction in post-neonatal injury death.

Conclusion

This evaluation found a number of estimated positive impacts for Family Start participants. Of these, most notable were the significant reductions in mortality within the child's first year of life. There were also positive estimated effects on PHO enrolment, immunisation rates, and Before School Check attendance. These positive impacts were also found for Māori and Pasifika children.[12] That said, these promising results were identified in the PSM analyses, with most unable to be replicated in the DiD analyses. These findings are broadly consistent with the findings in Vaithianathan et al. (2016).

While DiD is generally considered the stronger method because it controls for time-invariant unobserved factors, we faced several challenges when applying this method in this study. Most notably we were not able to accurately identify the target group (i.e., children that might participate in the programme). This and other limitations decreased the likelihood of detecting the effects of Family Start. Because a lack of evidence may not indicate an absence of an effect, we do not interpret the lack of significant findings using the DiD approach as evidence that Family Start is not improving outcomes for participants.

More broadly, this study demonstrates the limitations that quasi-experimental techniques such as PSM and DiD can have in evaluating social service initiatives, and that these methods may not always be able to provide the conclusive evidence on programme impacts being sought. While this study found many promising indicators of positive impacts of Family Start participation for children and their whānau, the analysis also reveals the limitations of using administrative data to assess a broad range of outcomes relevant to child and whānau wellbeing, particularly for non-Western conceptualisations of wellbeing. In addition, many of Family Start's short-, medium-, and long-term goals are not able to be assessed by the data available in the IDI.[13] These issues are consistent with some of the challenges discussed in Wilson et al. (2018)[14] and Matheson (2020).[15] Therefore, we recommend that the findings from this impact evaluation should be considered alongside

[12] Note that for Māori and Pasifika children, some of the reductions in mortality were only significant at the 10% level.

[13] For example, that family/whānau is supported by their community of interest (iwi/hapū/church).

[14] Wilson, M., Hyslop, D., Belgrave, M., Vette, M., and McMillen, M (2018). *Estimating the impact of Social Workers in Schools using linked administrative data*. Ministry of Social Development, Oranga Tamariki.

[15] Matheson, I (2020). *Oranga Tamariki Early Intervention: A synthesis of recent research and evaluations.*

findings from the qualitative process evaluation of Family Start, as summarised in the synthesis report.[16]

BACKGROUND

Family Start overview

Family Start is a voluntary home-visiting programme funded, monitored and overseen by Oranga Tamariki (previously the Ministry of Social Development, MSD). It supports whānau at risk of facing health and social challenges to realise better outcomes for their children. Whānau are typically enrolled in the programme before the birth or in the registered child's first year of life,[17] and may remain enrolled in the programme until they transition into school or until support is no longer needed.[18] To enrol in the programme, whānau must be experiencing/have experienced:[19]

- mental health issues
- addiction problems
- childhood history of abuse (for the parent/caregiver)
- care or protection history
- relationship problems (including family and whānau violence)
- parenting or child health and development issues, or
- young parenthood with additional challenges or needs.

Family Start is provided by a number of providers using the Family Start approach, where each provider delivers services to a particular area (which may include multiple Territorial Local Authorities, or TLAs). Referrals are made directly to Family Start providers by local social services and agencies, or through family or self-referral. Family Start providers then aim to contact the family within five days to organise an initial visit to confirm eligibility and willingness to engage in the programme.[20]

[16] Carter, M., & Cording, J. (2021). *Evaluation of the Family Start Programme: Synthesis of process and impact evaluation findings*. Oranga Tamariki—Ministry for Children.

[17] Children may be enrolled between ages one and two in exceptional circumstances; see Oranga Tamariki (2020). *Family Start programme manual*. https://orangatamariki.govt.nz/assets/Uploads/Support-for-families/Support-programmes/Family-Start/Family-Start-manual.pdf

[18] Support is no longer needed where families have "achieved their goals, the child's wellbeing is enhanced, parents' confidence is increased, and living circumstances are improved" (p2, Oranga Tamariki, n.d., *Family Start referral guide*. https://www.orangatamariki.govt.nz/assets/Uploads/Support-for-families/Support-programmes/Family-Start/Family-Start-referral-guide.pdf)

[19] To be accepted into the programme, applicants are required to satisfy one or more of these conditions. Alternatively, they can be accepted to the programme by satisfying a combination of other conditions. These include sudden unexplained death indicators (e.g. smoking during pregnancy), lack of positive support networks, multiple births or short inter-pregnancy intervals, criminal justice involvement, financial and material resource difficulties, frequent change of address or housing issues, and parent educational difficulties.

[20] Oranga Tamariki (n.d.). *Family Start referral guide.*

Once a child is enrolled with Family Start, home visits are undertaken by a Family Start worker. Service delivery is guided by a programme manual developed by Oranga Tamariki.[21] The manual specifies core service delivery components (the Parenting Resource, Strengths & Needs Assessments (SNA), Child Safety Tools (CST) and Child Family Plans (CFP)). The delivery of these core components follows a cyclic process.

Family Start operates from a child-centred, strengths-based approach, to encourage whānau to explore options available for managing problems and difficulties encountered when raising young children. The programme is also designed to help parents enjoy raising their children in a way which promotes healthy outcomes. Family Start facilitates outcomes by:

- encouraging whānau to build strong bonds between the parent and child
- developing whānau safety awareness
- teaching whānau about healthy lifestyle choices and child nutrition, health visits and immunisations
- developing parenting confidence.

Family Start workers respond to the unique needs of the whānau they are working with, and ultimately how service delivery is undertaken depends on the needs of the whānau. Ultimately, Family Start aims to support vulnerable children and reduce maltreatment.

The Theory of Change for the Family Start programme, including short-, medium-, and long-term outcomes, is presented on the following page in Figure 1.

Family Start history

Family Start was originally piloted in 1998 in the Rotorua, Waitakere and Whangārei Districts.[22] Following the initial pilot, the programme was introduced to an additional 13 Territorial Local Authorities (TLAs) during the period 1999-2000. Family Start later became available in an additional 14 TLAs during the period 2005-2007 (referred to in this report as 'the 2005-2007 Family Start expansion'). For the next nine years Family Start did not expand further, although a number of changes to the programme content and structure occurred throughout this time.

The 2017 Budget allocated an additional $28 million over four years to expand Family Start, with a total of $47 million per annum to be spent on the programme. The nationwide expansion, which began in 2017, has resulted in the provision of Family Start in an additional 37 TLAs to date. Table 2 in the Methodology section summarises the year in which Family Start became available in each TLA.

---

[21] Oranga Tamariki (2020). *Family Start programme manual.*

[22] Of note, Early Start, a similar early-childhood intervention programme, has operated in Christchurch from the mid-1990s to the present day. Early Start was excluded from the current evaluation, and analyses therefore do not include children from the Christchurch City TLA.

Figure 1. Family Start Theory of Change model (from Oranga Tamariki [2020]. *Family Start Programme Manual*, p. 14)

Theoretical framework

- Family Partnership model
- Ecological model
- Strengths-based practice model
- Attachment theory
- Neuro-developmental research
- Tikanga Māori

The Family Start programme provides a voluntary intensive home visiting service that promotes safe parent–child relationship and maternal, infant and early childhood wellbeing in high-need families.

Core quality components (inputs/resources)

- Qualified and professional Family Start workers
- Quality Family Start worker–client relationships
- Reflective clinical supervision
- Effective data management and support
- Community outreach and cross-agency co-ordination
- Adequate and sustained funding
- Parent education

Core service components (outputs/activities)

- Regular and on-going home visits
- Supports parents' understanding of child development and positive parenting practices
- Identification of issues impacting on the health and wellbeing of the child (e.g. alcohol and drugs, family violence)
- Identification of social connections to community/whānau/hapū/iwi
- Setting of specific, measurable, attainable, realistic, timely (SMART) goals to achieve families' aspirations

On-going support and continuous quality improvement

Assumptions

- Children's development and health will improve if parents are supported to access health and education services.
- When parents understand child development their children are more likely to have improved circumstances.
- Children's wellbeing will improve if parents understand the detrimental effect that family violence and addiction can have on their children.
- If the issues of family violence, post-natal depression and addiction are not addressed in families, children are less likely to become well-adjusted adults.
- Voluntary programmes work best with families that are willing to make positive changes in their lives.

Short-term outcomes

1. Child's health and safety

- Children's health and safety issues are identified early and addressed.
- Parents increase confidence in their ability to keep their children safe.
- Parents increase knowledge of and confidence in ability to positively engage with Well Child/Tamariki Ora and primary health organisations (PHOs).
- Address family violence and alcohol and drug misuse that impact on the child (access to specialist community services).
- Identification of and access to services for mothers with post-natal depression.
- Child/family enrolled with PHO.
- Child enrolled with Well Child/Tamariki Ora and oral health provider.
- Child immunisations are up-to-date.

2. Parenting skills and practices

- Parents increase knowledge of child development.
- Parents increase knowledge of and confidence in their ability to provide warm, sensitive, consistent and competent child care.

3. Participating and connected

- Families increase knowledge of other supportive family members and services and the confidence to access them.
- Parents have access to community resources to assist with child development.

4. Personal and family circumstances

- Families have a sense of hope and increase their knowledge of future options, including educational opportunities and employment.
- Family/whānau is supported so they can culturally identify with their community of interest (iwi/hapū/whānau/church).
- Parents increase knowledge and confidence to connect with a supportive community.

Medium-term outcomes

1. Vulnerable children/tamariki live in a safe environment

- Family members are receiving treatment for family violence and alcohol and drug abuse services.
- Mothers with post-natal depression are receiving treatment.

2. Family/whānau demonstrate improved parenting skills and practices

- Parents routinely spend time interacting in a nurturing and positive manner with their children.
- Parents have improved parenting attitudes and exhibit positive parenting behaviour.
- Parents demonstrate knowledge of their children's developmental abilities and emerging skills and stages.
- Parents have increased ability to accomplish their goals.

3. Families/children have connection to formal and Informal supports; whānau; hapū; iwi

- Family/whānau is supported by their community of interest (iwi/hapū/whānau/church).
- Parents have connections and are active in a supportive community.
- Child is participating in quality early childhood education.

4. Family/whānau demonstrates improvements in their personal and family circumstance.

- Parents are more optimistic about the future (have a sense of hope).
- Parents have strong relationships that are nurturing, violence-free and addiction-free.

Long term outcomes

- Reduction in child maltreatment
- Children are physically and mentally healthy
- Children/tamariki/families/whānau are healthy and resilient

Evaluation outcomes measures

- Population outcomes
- Results-based accountability (RBA) outcomes

Previous evaluations of Family Start

In 2005, an evaluation of the impact of Family Start involving four programme sites (West Auckland, Hamilton, Whakatane, and Nelson) was conducted.[23] The evaluation found that Family Start appeared to improve rates of parenting knowledge, caregiver participation in education and employment, and access to a child health worker. However, findings suggested that the programme did not increase rates of breastfeeding or child immunisation, nor decrease caregiver smoking. Notably, this evaluation did not include control or comparison groups, but instead measured differences in outcomes across a 12-month period. It was concluded that further impact evaluations were required to assess the effect of Family Start on longer-term outcomes for children and their family or whānau.

In 2009, an independent assessment of the Family Start and Early Start programmes was conducted, including a review of Family Start monitoring data and interviews with providers, enrolled programme participants, and other stakeholders.[24] The review found that Family Start appeared to have a positive effect on breastfeeding, Early Childhood Education enrolment, immunisation, and completed Well Child visits, however as the analysis was based on a comparison with national-level averages rather than a more precise quasi-experimental design these results are of limited plausibility. The effectiveness of the programme was found to vary across providers. The review concluded that Family Start had "considerable potential" (p. 2) to effect positive outcomes for children, but that consistency in programme content and delivery was required, supported by changes in the contracting structure. This led to the standardisation of contracted programme elements across providers in 2011-2012.

MSD commissioned Vaithianathan et al. (2016)[25] to carry out a quantitative analysis of the impact of Family Start participation/availability. The authors analysed the impact of Family Start using two quasi-experimental methods:

- Propensity Score Matching (PSM)-based individual-level analysis of the impact of Family Start participation for children and their whānau
- Difference-in-Difference (DiD)-based group-level analysis of the impact of being born in an area where Family Start is available, for children whose mothers received a benefit during the first thirteen weeks of their lives.

Key findings included a statistically significant reduction in the probability of post-neonatal mortality during the first year for children who participated in Family Start. Notably, this reduction was estimated using both DiD and PSM methods. In addition, the PSM estimated statistically significant reductions in the year one post-neonatal SUDI and injury-related death. These findings

[23] Centre for Child and Family Policy Research (2005). *Outcome/impact evaluation of Family Start: Final report.* Report prepared for the Centre for Social Research and Evaluation, Ministry of Social Development. https://www.msd.govt.nz/about-msd-and-our-work/publications-resources/evaluation/outcome-impact-family-start/index.html

[24] Cribb, J. (2009). *Review of Family Start and Early Start.* Report prepared for the Minister of Social Development and Employment. http://msd.govt.nz/documents/about-msd-and-our-work/publications-resources/official-information-responses/2019/august/20190828-review-of-family-start-and-early-start-unpublished-report-prepared-for-msd-research-to-inform-improvements-to-increase-family.pdf

[25] Vaithianathan, R., Wilson, M., Maloney, T., & Baird, S. (2016). *The Impact of the Family Start home visiting programme on outcomes for mothers and children: A quasi-experimental study.* Ministry of Social Development.

were of considerable policy significance, representing an objective indicator of the positive impact of Family Start on the lives of children.

The efficacy of Family Start for selected sub-groups of participants was assessed as an extension to the original 2016 impact evaluation (Vaithianathan et al., 2017).[26] These additional analyses found that there were significant reductions in post-neonatal mortality for several sub-groups, including:

- children with teenage or non-teenage mothers
- children in families with and without previous contact with CYF
- Māori children enrolled with kaupapa Māori and mainstream providers.

The extension analyses also suggested that positive outcomes for Māori children, including increased enrolment with a Primary Health Organisation at age one and immunisation rates, were more consistently observed for those children enrolled with a kaupapa Māori, rather than mainstream, organisation.

A subsequent study that braided the findings with Māori knowledge streams suggested that one possible explanation for the positive results for Māori was the role a cultural overlay to the parenting programme used by Family Start – Āhuru Mōwai and Born to Learn – played in engaging parents of Māori infants in the programme and supporting positive parenting. Another was Māori organisations combining funding streams to deliver a more holistic service to whānau.[27]

Current Family Start evaluation

In the context of the nationwide expansion of Family Start, Oranga Tamariki commissioned a process and impact evaluation intended to build on previous studies that have examined the impact of the programme (i.e. Vaithianathan et al., 2016). The evaluation comprises a quantitative impact evaluation (the current report), a qualitative process evaluation,[28] and an overall synthesis report.[29]

This impact evaluation's purpose was to build on the approach in Vaithianathan et al. (2016), applying a similar approach to assess the impact the programme is having for children and their whānau in terms of specific outcomes (e.g., altering the probability that a child will appear in Oranga Tamariki initial intake data during their first 12 months of life). In addition, higher-level evaluative judgements are made for three outcome groupings: impacts on child protection-related outcomes, impacts on post-neonatal mortality, and impacts on other health- and education-related outcomes.

[26] Vaithianathan, R., Maloney, T., Wilson, M., & Joyce, S. (2017). *Family Start impact study: Selected extensions.* Ministry of Social Development. https://www.msd.govt.nz/documents/about-msd-and-our-work/publications-resources/evaluation/family-start-outcomes-study/family-start-impact-study-selected-extensions.pdf

[27] Cram, F., Vette, M., Wilson, M., Vaithianathan, R., Maloney, T., & Baird, S. (2018). *He awa whiria—braided rivers: Understanding the outcomes from Family Start for Māori*. Evaluation Matters, 165-207. https://www.nzcer.org.nz/system/files/journals/evaluation-maters/downloads/Online_Articles_txt_Cram_FA.pdf

[28] Oranga Tamariki & Allen + Clarke (2020). *Evaluation of the Family Start Programme: Report on findings of the process evaluation.* Oranga Tamariki—Ministry for Children. https://orangatamariki.govt.nz/assets/Uploads/About-us/Research/Latest-research/Family-Start/Evaluation-of-the-Family-Start-programme.pdf

[29] Carter, M., & Cording, J. (2021). *Evaluation of the Family Start Programme: Synthesis of process and impact evaluation findings.* Oranga Tamariki—Ministry for Children.

The impact evaluation is intended, in conjunction with the process evaluation, to provide accountability for current investment, inform future investments in the programme, and support continuous improvement and learning to optimise positive outcomes for children and their whānau.

Key evaluation questions

The key evaluation questions (KEQs) the broader Family Start evaluation sought to answer were:

1. How well is Family Start delivering its service for vulnerable children and their whānau?
2. To what extent is Family Start achieving programme outcomes and impacts for vulnerable children and their whānau?
3. How can Family Start be optimised to ensure positive outcomes for children and their whānau?

The impact evaluation focused on answering KEQ2.[30]

Performance indicators for impact evaluation

The evaluation team developed a criteria framework to inform assessments of the Family Start programme's impact. This included articulation of desired achievements for the programme, as well as specific performance indicators. The desired achievements for the programme and agreed performance indicators identified in the Evaluation Plan (which are all elements of KEQ2) are provided in Appendix A Table A1.

Following the finalisation of the Evaluation Plan, further collaboration with Oranga Tamariki's Evaluation Advisory Group (EAG) helped refine and prioritise the specific outcomes that could most meaningfully be explored in the current evaluation. Of particular note, it was agreed that a number of outcomes identified relating to maternal wellbeing were difficult to plausibly link with Family Start and these outcomes were not further explored. The potential for future evaluations to address this issue using maternal-level matching is noted in the Discussion section.

During the analysis we modified the outcome groupings set out in the Evaluation Plan (see Appendix A); results are instead reported and presented using the groupings set out below in Table 1. The original outcome groupings were modified for the following reasons:

- post-neonatal mortality is a particularly significant topic and warrants separate discussion
- other results relating to child protection can sensibly be grouped together
- some of the original outcome groupings overlapped in content, making it efficient to combine them.

Appendix A Table A2 presents the rubric used in the Results section of this report to assess to what extent the desired outcomes of Family Start have been met.

[30] The process evaluation focused on answering KEQ1. As an exploratory and descriptive question, KEQ3 is focused on identifying learnings and improvements, and guided the development of evaluation recommendations provided in the overall evaluation synthesis report.

Table 1. Outcome groupings

Desired achievement	Specific outcomes
Family Start participation has a positive impact on post-neonatal mortality	• Post-neonatal mortality • Post-neonatal injury death • Post-neonatal SIDS/SUDI
Family Start participation has a positive impact on other health and education-related outcomes	• Childhood immunisation • Enrolment with a Primary Health Organisation (PHO) • Enrolment in early childhood education (ECE) services • Attendance at the Before School Check (B4SC) • Children identified through the B4SC as within healthy BMI range • Prevalence of 'significant issues' and 'non-significant issues' identified at the B4SC
Family Start participation has a positive impact on child protection outcomes	• Oranga Tamariki (previously Child Youth and Family) Intakes, Assessments, and Placements in Care[31] • Child maltreatment or long-bone fracture hospitalisations

[31] Intakes and Assessments are analogous to the Initial and Core Assessment phases, respectively.

METHODOLOGY

Bridging Cultural Perspectives

The current evaluation was informed by the Bridging Cultural Perspectives approach[32] which comprises two models: He Awa Whiria (Braided Rivers) and Negotiated Spaces.

He Awa Whiria provides the framework for knowledge creation. It provides two separate streams of knowledge – Māori and Pākehā – each stream of equal strength, with information about what is valued, and to what degree. For the Family Start evaluation, a Pasifika knowledge stream[33] was also woven into the evaluation process to produce findings based in each of the three knowledge streams: Māori, Pasifika and Pākehā.

Negotiated Spaces provides the dialogue tool for exchanging knowledge across the streams. Implicit to the Negotiated Space is balancing the desire to uphold distinctive cultural knowledge spaces with an openness to innovation and change.

In implementing the evaluation, reviewers representing the three knowledge streams (Māori, Pasifika, Pākehā) worked together to assist with the interpretation of the findings at an aggregate level, and through the lens of each worldview.

Sample creation

The study population was generated using data from Statistics New Zealand's Integrated Data Infrastructure (IDI).[34] The initial sample included 796,719 distinct children born between January 2003 and December 2015, who had a valid date of birth and sex records, had at least one recorded ethnicity, could be linked with a female parent ('mother'),[35] linked to the IDI spine,[36] and had a 'person indicator' flag. Next, we removed children without an address at birth,[37] reducing the sample size to 767,115. Removals were largely from early birth cohorts, with more than half of children removed being born between 2003 and 2005. Finally, we restricted the sample to only include children that were linked to at least one Ministry of Health dataset, resulting in a final sample with 765,048 observations of distinct children.

---

[32] Superu (2018). *Bridging Cultural Perspectives.* http://www.superu.govt.nz/sites/default/files/Bridging%20Cultural%20Perspectives%20FINAL.pdf

[33] The Pasifika knowledge stream comprises the cultural beliefs and world views of at least seven ethnic groups. These worldviews and beliefs can be considered together through the concept of Fofola e fala kae talanoa e kāinga, a metaphor of which one underlying meaning is an invitation to family members to come together and talanoa – to talk (Ministry of Social Development, 2012. *Nga vaka o kāiga tapu. A Pacific Conceptual Framework to address family violence in New Zealand.* Ministry of Social Development). Multiple mats can be rolled out simultaneously relating to both different ethnic groups and different parts of the evaluation.

[34] For more information, see: https://www.stats.govt.nz/integrated-data/integrated-data-infrastructure

[35] Reference children who could not be linked with a mother were excluded because a number of key characteristics – such as address – were assigned via mothers.

[36] This restriction ensured that children could be linked across datasets (e.g., health or education data).

[37] This was based on mothers' addresses, following the assumption that children live with their mothers.

Demographic information (birth year and month, sex, and ethnicity) was sourced from the IDI personal details table,[38] while address information was sourced from the address notification table.[39] Note that, as in Vaithianathan et al. (2016), we excluded births from Christchurch City as this TLA was served by the Early Start programme.[40]

Children's potential to participate in Family Start was assessed based on the official availability of Family Start at the time of the child's birth in the TLA that the birth meshblock was located in. We obtained information about the official availability of Family Start based on data from Vaithianathan et al. (2016) and the Family Start team within Oranga Tamariki (relating to the nationwide expansion which began in 2016).Table 2 presents the expansion of Family Start up until December 2017 by year.[41]

Table 2. Family Start availability

Year	Territorial Authority
1995	Christchurch City (Early Start)
1998	Rotorua, Waitakere, Whangarei (pilot)
1999/2000	Dunedin, Far North, Gisborne, Hamilton, Hastings, Horowhenua, Invercargill, Kawerau, Masterton, Nelson, Porirua, Whanganui, Whakatane (first expansion)
2005-2007	Auckland, Opotiki, Lower Hutt, Manakau, Napier, Taupo, Waikato, Papakura, Buller, Grey, Hauraki, Wairoa, South Waikato, Ruapehu ('2005-2007 expansion')
2017+	Family Start has continued its nationwide expansion, with evidence of increased Family Start participation in all remaining TLAs

Reference child characteristics and outcomes

The estimation methods used in this evaluation required the total sample to be linked with other datasets in order to fully characterise reference children (including characteristics relating to parental mental health, benefit receipt and criminal convictions prior to the reference child's birth). Table C2 (Appendix C) provides a detailed description of all reference child characteristics that were derived, and how they were used in the different estimation methods.

[38] This table provides Statistics New Zealand's 'best estimate' for each of these characteristics by triangulating and prioritising information about a given individual from multiple sources.

[39] This table prioritises address records from multiple sources using a set of business rules to provide a 'best estimate' meshblock-level address for individuals at any given point in time.

[40] Note that because we used 2010 TLA definitions which merged the Banks Peninsula TLA into Christchurch City, we also excluded children born in the Banks Peninsula TLA from our analyses; Vaithianathan et al. (2016) did not exclude children born in the Banks Peninsula TLA. Because of the small size of the Banks Peninsula TLA, we do not anticipate this had a measurable impact on results. The 2010 TLA classifications were used in the current evaluation to capture participation within the different Auckland TLAs prior to its amalgamation into one TLA in 2013.

[41] More detailed information about the exact year and month when Family Start became available in each TLA was obtained and informed dataset construction and analysis but is not presented here for reasons of concision.

The total sample was also linked with the mortality, child protection, and health and education outcomes listed in Table 1. A detailed description of how each outcome was defined and created is provided in Appendix C.

Family Start data in the IDI

Family Start child participant data (and primary caregiver data) are available in the IDI, sourced from Oranga Tamariki's FS-NET. Introduced in late 2008, FS-NET is the Family Start data management system, capturing different aspects of the programme and its participants. FS-NET data available in the IDI contain limited details such as Oranga Tamariki-sourced demographic data and programme participation start and end dates. Analysis of the FS-NET participant data is presented in Appendix B.

Children were linked to Family Start data using a unique identifier, *snz_uid*. Children who appear in the Family Start data but who could not be linked to the aforementioned initial sample extracted from the IDI spine are not included in any of the analyses presented in this report. A brief analysis of the demographic characteristics recorded in the Family Start data of children who could not be linked with the IDI spine did not reveal any differences between those who were linked and those not linked, except that children who were identified as Pasifika in the Family Start data were slightly less likely to be linked to the IDI spine. We speculate that this may reflect linkage issues caused by variation in the spelling of first or last names, which may be more common in the administrative records of Pasifika people. This difference was relatively minor and does not introduce any obvious bias to the analyses presented in this report.

Table 3 presents some key characteristics of Family Start children (and their parents) who were born in the period 2009 to 2015.[42] It also provides a comparison with children born in the same TLAs who did not participate in Family Start. Characteristics have been presented for all children (i.e. any ethnicity), and for Māori and Pasifika children separately.

Close to 60% of children in the participant sample used for the PSM analysis identified as Māori, followed by just under 30% who identified as Pasifika (including 18.4% of the total sample who identified as Pasifika but not Māori). Just under half of the participant sample was female. Examples demonstrating the high need profile of Family Start children and their whānau include the proportion of children born in NZ Dep 9 or 10 area units (64.3% total, 69.5% Māori, 74.7% Pasifika), high proportion of mothers supported by benefits within three months of birth, differences in mother relationship and age at child's birth, prior parental interactions with the Ministry of Health, Department of Corrections, and social services, and a greater number of total risk factors.

[42] We restrict the analysis to observations recorded from 2009 onwards due to poorer data quality of earlier engagements.

Table 3. Family Start participants vs non-participants (2009-2015, TLAs where Family Start was available)

	Total		Māori		Pasifika	
Characteristic	Participants	Non-participants	Participants	Non-participants	Participants	Non-participants
Children	16,761	229,437	9,972	71,007	4,980	49,353
Māori	0.595	0.309	1.000	1.000	0.382	0.299
Pasifika	0.297	0.215	0.191	0.208	1.0	1.0
Female	0.481	0.486	0.483	0.485	0.477	0.485
Child had a low birth weight	0.099	0.059	0.102	0.065	0.080	0.050
Child was born in NZDEP 9-10 area	0.643	0.348	0.695	0.521	0.747	0.614
Child was born in a major urban area	0.501	0.631	0.402	0.463	0.767	0.825
Child is their mother's first born	0.428	0.438	0.397	0.345	0.415	0.361
Mother						
Single at the birth of the child	0.669	0.270	0.766	0.515	0.632	0.430
Supported by benefit within 3 months of child's birth	0.698	0.266	0.770	0.487	0.677	0.427
Under 18 when child was born	0.065	0.010	0.080	0.022	0.066	0.017
Under 20 when child was born	0.224	0.058	0.256	0.116	0.224	0.097
Under 25 when child was born	0.566	0.250	0.612	0.415	0.556	0.382
Smoked at the time of the child's birth	0.387	0.138	0.486	0.303	0.270	0.159
Supported by benefits for 3 or more of the last 5 years	0.312	0.119	0.374	0.244	0.263	0.175
Received mental health services or prescription in the 5 years before the child's birth	0.333	0.107	0.346	0.159	0.230	0.091
Had a child protection placement before age 18	0.121	0.024	0.143	0.050	0.093	0.031
Had a child protection notification before age 18	0.468	0.140	0.548	0.284	0.388	0.186
Sibling had a child protection placement during the 5 years prior to child's birth*	0.050	0.008	0.060	0.019	0.035	0.010
Sibling had a child protection notification during the 5 years prior to child's birth*	0.341	0.106	0.395	0.211	0.311	0.154
Served a community service sentence during the 5 years prior to child's birth	0.160	0.040	0.203	0.096	0.104	0.043
Served a custodial sentence during the 5 years prior to child's birth	0.035	0.005	0.047	0.014	0.024	0.005
Father						
Recorded on birth certificate and linked to IDI spine	0.853	0.946	0.844	0.903	0.862	0.916
Received an incapacity benefit due to substance abuse in 5 years prior to child's birth	0.036	0.009	0.041	0.020	0.017	0.007
Received an incapacity benefit due to mental health in 5 years prior to child's birth	0.081	0.025	0.081	0.044	0.037	0.016
Had a child protection placement before age 18	0.085	0.021	0.109	0.048	0.061	0.029

Had a child protection notification before age 18	0.302	0.106	0.372	0.229	0.233	0.133
Served a community service sentence during the 5 years prior to child's birth	0.325	0.110	0.392	0.232	0.260	0.150
Served a custodial sentence during the 5 years prior to child's birth	0.152	0.039	0.195	0.092	0.104	0.049
Average number of risk factors (child)	5.4	2.1	6.2	3.6	4.8	2.9

Notes: Counts and have been randomly rounded, and shares were based on randomly rounded counts using Stats NZ's random rounding 3 regulations. For full list of variables, see Appendix C. 'Sibling' refers to all other children linked to the child's mother using DIA data; this can include birth siblings who no longer live in the same household as the mother or participant child, and exclude step-siblings or other unrelated children who live in the same household as the mother or participant child.

Period of analysis

We initially included births from January 2000 to December 2017, but later shortened the time period used in the analyses. Births between 2000 and 2002 were excluded due to concerns about poor data quality. Births between 2016 and 2017 were also excluded due to a lack of outcome data available in the IDI at the time of analysis, and because of an observed rise in Family Start participation for children who were born in areas where Family Start was not officially available; this increased the probability of control samples containing children who had in fact participated in Family Start.

Note that when applying different estimation methods for analysing the data, different periods of time were used. This is discussed in the estimation section below.

Estimation

Overview and relationship to Vaithianathan et al. (2016)

Two quasi-experimental methods were used in the current evaluation: Propensity Score Matching (PSM) and Difference-in-Differences (DiD). These methods provide potentially complementary approaches with different strengths and weaknesses in the context of this evaluation of Family Start. Quasi-experimental methods were required for this evaluation because participation in Family Start is not randomly assigned. Participation is voluntary and based on a variety of eligibility criteria largely related to risk factors indicative of poor child outcomes.

The current analysis was heavily informed by Vaithianathan et al. (2016)'s evaluation and has benefited greatly from the generous provision of code by the authors. For the most part the outcome measures adopted are similar, although in some cases we have used modified variable definitions using IDI code developed by Oranga Tamariki actuaries in collaboration with Ernst & Young.

We built on the quasi-experimental approaches developed by Vaithianathan et al. (2016) and took advantage of the additional data now available in the IDI to test and refine their approach. The current impact evaluation extended the previous evaluation in the following ways:

- as a result of analysing observed outcomes across longer periods of time and therefore including more children in the research sample, analyses have greater statistical power to detect the impact of Family Start on rare outcomes such as mortality
- we modified the model specifications to increase confidence in the robustness of findings presented.

The results reported in the main body of this report have been estimated using this modified modelling approach, whereas Appendices D and E provide details on the extent to which we were able to replicate the findings presented in Vaithianathan et al. (2016) when using the same models and analysing the same time periods. For the most part, results appeared reassuringly similar when comparing PSM outcomes. DiD outcomes were also similar with the exception of the key DiD finding presented in Vaithianathan et al. (2016) of a statistically significant reduction in post-neonatal mortality in the first year of life. However, this outcome could be replicated when adding population weights to the model. In addition, some specifications that included weights also estimated a statistically significant reduction in post-neonatal injury death, which was not estimated in the original analysis. Complicating this finding, we note in Appendix E that relatively minor variations in the period analysed as part of the DiD replication attempt (increasing the length of the period analysed by including data from 2003q1 onwards rather than 2004q3 onwards) resulted in the estimation of non-significant mortality impacts. We have interspersed commentary on the consistency of our current results with the original findings throughout the results section of this report.

Propensity Score Matching (PSM)

Propensity Score Matching (PSM) is a statistical approach used to create comparable treatment and control groups in situations where randomised allocation to treatment and controls groups (as required for Randomised Control Trials, RCTs) is not possible for ethical or pragmatic reasons, or when an initiative is set up without a control group. In this way, PSM is used as an approximation of an RCT design in quasi-experimental research.

In a standard PSM model, this is achieved by first identifying which factors are most strongly associated with individuals participating in the programme being evaluated (i.e. identifying relevant characteristics of individuals in the treatment group based on theory and previous research). Using these characteristics, a logistic regression is estimated for calculating the probability of participating in the programme – this is known as the 'propensity score'. A propensity score is calculated for all individuals who participated in the programme ('treated individuals'), as well as for all untreated individuals who could potentially be used as controls in the evaluation ('control individuals'). Treated individuals are then matched with control individuals who have very similar propensity scores; this means that theoretically, each treated individual is matched with one or more control individuals who had an equal probability of participating in the programme.

Mean outcomes for the matched treatment and control groups are then compared using standard analytical techniques such as regression or *t*-tests. The underlying assumption is that once the characteristics predicting participation are controlled for by PSM, the difference in mean outcomes between groups is determined by their participation (or absence of participation) in the programme.

The validity of PSM relies on the assumption that there are no unmeasured characteristics that predict participation and affect the outcomes of interest. If this assumption is not met, the estimated impacts may be biased. Unmeasured predictors that could result in either positive or negative bias can be imagined in the case of Family Start. Participation in Family Start is voluntary, and as such unmeasured predictors of participation could include optimism, positive motivation and/or resilience, all of which are likely to also positively bias the outcomes measured. Conversely, Family Start participation often involves referrals from social services. The fact that a child's family or whānau has come to the attention of social services suggests that they may be facing additional challenges that have not been directly measured (and therefore matched between groups); these additional challenges could negatively bias the outcomes measured.

The variety of referral channels available for Family Start also means that it is difficult to reliably capture relevant predictors of participation for each unique referral pathway, particularly for referral criteria that are not directly captured in the IDI (e.g., parenting issues, relationship problems).

Approach to PSM taken in the current study

The PSM modelling approach taken in the current evaluation comprised two stages. Stage one involved estimating a propensity score model for children who lived in areas where Family Start was available at the time of their birth. The propensity score model was generated using a standard logit regression model, as shown below. The dependent variable Y indicates Family Start participation, the independent variables $(Z_1, \dots, Z_j)$ included in the model are largely the same predictors of programme participation identified by Vaithianathan et al. (2016; see Appendix C of this report for a complete list of variables included)[43] and f represents the cumulative logistic distribution function:

$$\Pr\left(Y = 1 \middle| Z_1, \ldots, Z_j\right) = f(\beta_0 + \beta_1 Z_1 + \cdots + \beta_j Z_j)$$

Following the initial stage one regression, the probability of participation was estimated for all children in the dataset.

Stage two involved matching participant children with suitable non-participant children (i.e., 'controls') based on propensity scores and the following exact matching criteria:

- whether the child's mother was supported by a main benefit within the first 3 months following birth, combined with mother's age (three categories: mother under 20, mother 20 and over and received benefit during first three months, mother 20 and over and did not receive a benefit during first three months)
- whether the child lived in a neighbourhood at birth that was in NZDep 9 or 10 (the most deprived quintile)
- the ethnicity of the child (Māori, Pasifika if not Māori, Other)
- urban location combined with Pasifika ethnicity (all Pasifika children coded as living in urban location to avoid matching issues due to the small proportion of Pasifika children born in non-urban locations).),
- birth year.

Controls were drawn from TLAs where Family Start *was* available (rather than those where it was not available, which was the approach in the 2016 evaluation). The benefit of this modification was that the geographic profile of participants and matched controls were very similar (i.e., this approach controlled for any regional differences in outcomes). Exact matching ensured that key characteristics of matched treatments and controls were identical and in practice the two groups were always very similar with respect to other observed characteristics when compared. This gave us confidence that the results estimated were not a consequence of observable differences between the groups. However, we cannot dismiss the possibility that unmeasured participation-related biases impacted results.

The second stage of the PSM analysis was implemented using the STATA command *psmatch2*,[44] with the Average Treatment effect on the Treated (ATT) estimated.

[43] The predictors of participation were determined at the time the reference child was born, for example, whether or not any siblings had contact with child-protection services before the child's birth. The only exception to this was whether the child was supported by a main benefit within the 13 months following birth. In addition, one notable difference is that we included location (TLA at birth) in the propensity model.

[44] We used *psmatch2* rather than the command available within STATA (*teffects nnmatch*) used in the 2016 evaluation. The exact matching was implemented in *psmatch2* by creating a match class variable (with took values 1 to 378 denoting the combination of the exact match criteria) and adding the propensity score to it. Using *psmatch2* allowed us to implement a calliper (set equal to 0.2*standard deviation of the estimated propensity score, which was 0.0190 in the main analysis) and to match individuals with up to five nearest neighbours rather than a single neighbour. Callipers set the maximum difference in propensity scores allowable for a potential match between a treated individual and a control individual to be made. Setting a calliper means that in cases where no close match exists, the participant is unmatched (and hence not included in the analysis). We used the Abadie-Imbens standard error option available in *psmatch2*. Because some of the outcomes were very rare events, we also calculated confidence intervals based on boot-strapping methods (1,000 resamples prior to propensity score estimation). This produced very similar results, although the estimated impact on post-neonatal SUDI was no longer significant at the p<.05 level.

Mortality extensions

To improve estimates of the impact of Family Start on child mortality, we modified the matching approach outlined above. Since most mothers join the programme after the birth of their child, the mortality impacts are biased by the existence of 'competing risks'. That is, mortality rates amongst those enrolled will necessarily be lower than mortality rates measured from birth (all other things being equal), because in order to enrol at a given age a child would need to have survived up to that age. This is exacerbated in the first year as, for all children, mortality risk is greatest in the months following birth, and even though the mortality measures used exclude the neonatal period, risks are greater in months two and three than month 12.

In Vaithianathan et al. (2016) this concern was addressed by estimating year one mortality outcomes for the sub-sample of children that enrolled in Family Start before four weeks of age (around one-fifth of participants), finding similar results to the main analyses.[45] We instead implemented a modified matching approach to address this issue, given we were interested in examining mortality outcomes for all participants and not just those who enrolled pre-birth (which even if significant, represents a sub-sample of the entire participant group and may not reflect the experience of all children).

Our modified matching process was implemented by creating 25 records for each potential control child (with month of age taking the value 0 to 24 months), deleting all month records following the month of death, and exact matching on the age (in months) the participant child entered the programme. The process proved relatively computationally demanding, which meant that we did not extend this approach to non-mortality outcomes.

In retrospect, we would have preferred to match children based on characteristics at the time they started Family Start (and the equivalent ages for non-participants), rather than characteristics at birth. This would have required us to create multiple records for each potential control child by month of age (as above) and then to derive the matching variables as at that age. This approach would have enabled us to control for the child's contact with child protection services (and other characteristics) prior to starting Family Start (including post-birth), thereby improving the reliability of findings from analyses and reducing the potential of a reverse-causality bias in results. We did not adopt this approach because of the computational demands that it incurs and because of timing issues.

PSM sample

Our main analysis focuses on children born in the period 2009 – 2015.[46] We include all participants who lived in TLAs where Family Start operated during this period.[47] Table 3 describes the characteristics of participants in our study population and compares these to those of non-participants.

---

[45] These results were not included in the 2016 report, and it is not clear whether the similarity was in terms of the estimate and/or statistical significance.

[46] In 2016 Family Start began to expand and was increasingly available in all areas. Although our PSM approach selected controls from TLAs where Family Start was available, restricting to this time period enabled us to evaluate the impact of selecting controls from TLAs where Family Start was not available (as Vaithianathan et al. (2016) did).

[47] Vaithianathan et al (2016) focussed on participants in TLAs where Family Start first became available during 2005 – 2007. They noted that in the TLAs where Family Start was operating prior to 2005, the programme operated under different contracting rules. Because Oranga Tamariki are interested in understanding the impact of the programme, we included participants from *all* TLAs where Family Start operated during 2009-2015.

As described in the previous section, participants were matched with up to five control individuals, based on the being an exact match on a small number of key characteristics and the propensity score calculated for each child. Only 306 out of the 16,764 children (1.8%) in the participant sample were not able to be matched to an appropriate control and were therefore not included in the final analyses. The total number of controls selected was 45,474 with a weight applied to adjust for the number of matches made in each case. A comparison of the characteristics of matched participants and matched (and weighted) controls is provided in Table D4 (Appendix D). The right-hand panel of the table shows that the differences in observed characteristics between the two groups were small.

Results from our primary model are presented in the main body of this report; results for other specifications are included in Appendix D, which also describes how our approach differs from that taken by Vaithianathan et al. (2016) and the impact this had on the main findings. Specifically, Table D5 contrasts estimated impacts from our primary model with those generated by matching participants with controls from areas where Family Start was not available, and prior to adjusting the matching approach used to estimate mortality impacts. Appendix D also includes our replication of Vaithianathan et al. (2016) analysis for the same period, using the same data and methods to the extent possible (Table D2).

Note that some outcomes are not observed for more recent cohorts. Mortality in the first year after birth was not observed for the 2015 cohort and mortality in the sixth year after birth was not observed for the 2011-2015 cohorts; in part, this reflects the time lag before Ministry of Health mortality data are finalised. Non-mortality outcomes in the sixth year were not observed for 2014 and 2015 birth cohorts.

Difference-in-Difference (DiD)

In a simple DiD analysis, outcomes for the treatment group (e.g., children born in TLAs with access to Family Start) are assessed before and after the introduction of a programme. Changes in outcomes between the two periods are measured relative to the changes in outcomes experienced by children in the control group (e.g., children born in TLAs without access to Family Start). If the only difference that affects outcomes over this period between the two groups was the introduction of Family Start in the treatment group, then these differences can be attributed to the programme. A key advantage of this approach is that the estimate measures change over time, meaning that differences between TLAs and cohort-specific features that are constant over time (i.e., 'time-invariant' differences) are controlled for.

DiD was used in Vaithianathan et al. (2016) as it addresses the concern that despite the matching process, the PSM estimates will still be biased due to unmatched time-invariant differences between participants and non-participants (e.g. motivation to participate in the programme).[48] Note that rather than focusing on participants, the DiD analysis examines outcomes at the area (TLA) level using a sub-sample of children that were expected to benefit from the programme (regardless of whether they participated, commonly termed 'intention to treat').

In this study, we use DiD for similar reasons. As in Vaithianathan et al. (2016), we focus on the 2005-2007 Family Start expansion as it fits with the DiD framework. Following this expansion, Family Start became available in 14 new TLAs (treatment group), while a remaining group of 35 TLAs did not have access to the programme during the entire study period (control group).

---

[48] This estimation strategy was termed Fixed Effects (Community Level) in Vaithianathan et al. (2016).

Note that for the results of the DiD analyses to be valid, any differences between the outcomes in Family Start and non-Family Start TLAs due to factors other than Family Start must remain constant over time. That is, any non-Family Start related time trends in the outcomes must be common across TLAs. This assumption is difficult to confirm, although the current evaluation conducted some tests to provide some information on this requirement.[49] Finally, since the uptake of Family Start is not universal in the treated TLAs, measures of the average area-level impact (i.e., the impact on all potential, rather than actual, participants) would be expected to result in smaller estimates of the impact of Family Start. Given 18% of children in the identified target group in treated TLAs actually participated in Family Start, we would expect estimated impacts to be much smaller than those obtained by PSM. If the impact of Family Start is only experienced by participants, the DiD estimates would need to be scaled up by a factor of five to provide comparable estimates to those in PSM.

Approach to DiD taken in the current evaluation

The DiD specification used in the current evaluation can be described using the following equation:

$$Y_{ikt} = \beta D_{kt} + \alpha X_{ikt} + \delta_k + \delta_t + \varepsilon_{ikt}$$

For a child i born in year-quarter t in TLA k, Y_{ikt} represents a given outcome (e.g., child was hospitalised with a maltreatment injury in the year after birth). D_{kt} is a binary variable which indicates whether Family Start was available in the child's birth TLA at year-quarter of birth. X_{ikt} is a set of child-related characteristics (controls). δ_k and δ_t are birth-TLA and birth year-quarter fixed effects, respectively. Finally, ε_{ikt} is a noise/random error term.

In this model, β is the parameter of interest, capturing the average impact of the availability of Family Start on outcome Y for a child born in a TLA where Family Start was available and who met the model inclusion criterion (see next sub-section for further discussion). In all models, standard errors are clustered by TLA; this is done to account for TLA-level characteristics which make children/ whānau living in a given TLA more similar to each other than to children/whānau living in other TLAs (e.g., variation in long-term unemployment rates between TLAs).

The specification used in the current evaluation used individual-level data to ensure that data from TLAs with more births had a greater influence on results. We also included an extensive set of control variables to explicitly address numerous potential sources of confounding (see Appendix C). Finally, we included TLAs that were defined as 'semi-treated' and excluded TLAs that had Family Start prior to 2004.[50]

The time period for the DiD analysis in the current evaluation included births from 2003 to 2015.[51] Including eight quarters in the pre-treatment period increased our confidence that any observed

---

[49] This assumption was tested using a joint statistical significance test. For more information, see Appendix E.

[50] Semi-treated TLAs were TLAs that were excluded in Vaithianathan et al. (2016) as the authors could not determine their treatment status. In Appendix E, we report how including these TLAs affects the DiD estimates. Note that the exclusion of TLAs that had Family Start prior to 2004 was done for consistency reasons with Vaithianathan et al (2016) and was not expected to materially affect the DiD estimates.

[51] Vaithianathan et al.'s (2016) evaluation included births from Q3 2004 – Q4 2011 in their DiD analyses. The current evaluation increased the DiD analysis period by seven quarters to extend the period of time prior to the expansion of Family Start, and four additional years following the expansion (2011 – 2015). Unfortunately, inclusion of early quarters resulted in some outcomes not being estimated in the DiD analyses due to lower data quality or availability in those early periods. On the other hand, including the additional four years (2012-2015) ensures that the estimates are less affected by the Global Financial Crisis, and are measured over a longer period. This allows measurement of second and sixth year outcomes; only first year outcomes were measured using DiD in Vaithianathan et al. (2016).

differences in outcomes are not a continuation of pre-treatment trends (i.e., the parallel trends assumption). This assumption was tested for each outcome using parallel trends (joint significance) tests; these are outlined when discussing the main results, with more detailed information in Appendix E.

Appendix E also details our attempt to replicate the sample and model from Vaithianathan et al. (2016), and how introducing changes to the model affected the results. Our results also show some sensitivity to the inclusion of weights and controls. We believe that weights accounting for the relative population size of each TLA at different periods, and the inclusion of controls and additional TLAs resulted in a more robust analysis.

Inclusion criteria / target group

It is vital to correctly identify the children that are likely to be affected by the programme when taking the area-level approach. However, as the criteria for acceptance to Family Start are highly diverse, many of the differences between participants and their unique situations leading to participation are not captured in the IDI. In their absence, the target group (termed 'target' henceforth) for the DiD analysis was identified by applying the first phase of the PSM model described above. We then calculated the propensity score for all children born between 2003 and 2015 from both treatment and control TLAs, restricting the DiD target sample to the children that recorded the highest 25% of propensity scores.[52] The rationale for using this approach was to focus our analysis on children that are similar to actual programme participants in terms of observed characteristics, rather than examining the entire child population. We found that between 2009 and 2015, the 'target' inclusion criterion captured 73% of Family Start participants.[53]

Based on our 'target' inclusion criterion, the DiD sample included 91,524 observations of children born between 2003 and 2015 in both 'phased-in' TLAs (treatment), and in TLAs for which the programme was not available (control).

DiD descriptive statistics

Table 4 presents the mean characteristics of our target group, as well as for all children born in the treatment and control TLAs. The table shows that the treatment group was relatively larger than the control group and included a greater share of: urban births (71% compared to 17% in the control); Pasifika children (46% compared to 13%); and births in meshblocks in a high deprivation quintile (77% compared to 38%). These differences are not surprising since the 2005-2007 expansion targeted TLAs that included high shares of births in highly deprived areas, prioritising cities and towns with the greatest number of such births (Vaithianathan et al., 2016). We expect that the greater share of Pasifika children reflects their concentration in large urban areas (South Auckland in particular) which were serviced by the programme.

Both the treatment and control groups show similar shares of children identified as Māori (about 60%), with low birth weight (8%), maternal benefit support (over 80%), and number of residential TLAs mothers lived in during the 6 months prior to the child's birth and 18 months after (1.6-1.7).[54] As expected, the table shows a greater share of participants in the treatment TLAs (18% compared

[52] Our key assumption was that the characteristics of participants and the target group did not change prior to 2009.

[53] The target group in Vaithianathan et al.'s (2016) was based on child's benefit status in the first 13 weeks of life. See Appendix E for a comparison of this original criterion with the criterion used in this evaluation.

[54] This captures the number of distinct TLAs resided by the mother from 6 months prior to the child's birth, and until the age of 1.5. This was used as a proxy for transience.

to 5%). The proportion of participants born in non-treatment TLAs was skewed by an increase in the number of participants that were born in non-treatment TLAs in 2015. This likely reflects children who were born in 2015 in TLAs where Family Start was not available, joining Family Start when it became available in their TLAs in 2016/17. More generally, we speculate that the proportion of participants born in non-treatment TLAs reflects whānau migration (to treatment TLAs) between the child's birth and enrolment, inaccuracies in the address information, and participation of children born outside Family Start service provision boundaries.

The relatively small proportion of Family Start participants in the treatment group (under one-fifth of the total treatment group) reflects the difficulties of using administrative data to accurately identify the target group for a programme with highly diverse entry criteria. Combined, these may compromise the ability of the DiD approach to accurately detect the effects of the Family Start programme, especially for rare events and when differences are small, or to underestimate the true differences.

With respect to the children's mothers, Table 4 shows that a greater share of children from the treatment group were born to a mother who was single at the time of the birth (79% compared with 73%), and who was supported by benefit income for more than three of the last five years prior to the child's birth (37% compared to 32%). Mothers of children from the control group were more likely to have mental health-related events recorded in the period up to five years prior to the child's birth (29% compared to 20%), and to have been notified to Oranga Tamariki (Initial Assessment phase / intake) before turning 18 (48% compared to 41%).

With respect to the children's fathers, Table 4 shows that children from the control group were about 4 percentage points more likely to be linked to a father. Because father-level outcomes for children not linked to a father are coded zero (i.e. absent) rather than missing, this means that comparison of father-level outcomes between the groups is likely to be biased.[55] With that in mind, the table shows that shares in the control group are 4-7 percentage points greater for characteristics such as serving community work sentences, being notified to Oranga Tamariki before turning 18, and receiving a (non-substance use) mental health-related incapacity benefit. Finally, there was a slightly greater share of children for whom older siblings were notified (initial assessment phase) to Oranga Tamariki in the five-year period prior to the child's birth.

[55] We could have examined these outcomes only for children that were linked to a father. We chose not to follow this option since it is possible that the likelihood of being linked to a father is not random and focusing on children with a link would also bias the differences.

Table 4. Mean characteristics in TLAs where FS was phased in during 2005-2007 (treatment) and never treated TLAs (control), Q1 2003 - Q4 2015

Characteristic	Total		Target group	
	Phase in (Treatment)	Never treated (Control)	Phase in (Treatment)	Never treated (Control)
Children	237,015	257,715	53,568	37,956
Family Start participant (2009-2015)	0.0584	0.0091	0.1810	0.0470
Māori	0.2737	0.2462	0.5922	0.6065
Pasifika	0.2848	0.0633	0.4639	0.1333
Mother under 20 at child's birth	0.0677	0.0535	0.2646	0.3044
Born in an urban area	0.8080	0.3456	0.7119	0.1723
Born in a meshblock with NZDEP quintile 9 or 10	0.3807	0.1197	0.7675	0.3847
Mother supported by benefit in child's first 3 months of life	0.2964	0.1906	0.8416	0.8052
Low birth weight	0.0626	0.0573	0.0791	0.0816
Meshblocks lived in during the 6 months prior to the child's birth and 18 month after	1.9750	1.9327	2.9091	3.2998
TLAs lived in during the 6 months prior to the child's birth and 18 month after	1.3124	1.3048	1.5870	1.7187
Mother				
Single at birth of child	0.2807	0.1819	0.7871	0.7328
Supported by benefit for more than 3 of last 5 years	0.1339	0.0828	0.3704	0.3173
Served a community work sentence in the 5 years prior to child's birth	0.0350	0.0293	0.1275	0.1563
Served a remand/custody sentence in the 5 years prior to child's birth	0.0059	0.0044	0.0243	0.0275
Recorded any mental health related events in the 5 years prior to child's birth	0.0713	0.0777	0.1990	0.2874
Notified to Oranga Tamariki before age 18	0.1198	0.1067	0.4138	0.4757
Father				
Recorded in the birth registration and linked to IDI spine	0.9332	0.9631	0.7963	0.8412
Served a community work sentence in the 5 years prior to child's birth	0.0958	0.0815	0.3008	0.3666
Served a remand/custody sentence in the 5 years prior to child's birth	0.0371	0.0297	0.1393	0.1627
Notified to Oranga Tamariki before age 18	0.0749	0.0564	0.2591	0.3280
Received incapacity benefit for substance abuse in the 5 years prior to child's birth	0.0071	0.0086	0.0255	0.0452
Received incapacity benefit for mental health non-substance abuse in the 5 years prior to child's birth	0.0191	0.0248	0.0557	0.1069
Siblings notified to Oranga Tamariki in the 5 years prior to child's birth	0.0957	0.0673	0.3203	0.2993

Notes: Figures are based on randomly rounded counts in accordance with Stats NZ's random rounding 3 regulations. Siblings refers to all other children linked to the child's mother. Values derived from less than 6 observations have been supressed. For full list of variables, see Appendix C.

Table 5 presents the proportion of children with records of the various outcomes used in the DiD analyses. These are presented for both the treatment and control groups before (Q1 2003 – Q2 2005; marked 'Pre' in the table) and following (Q3 2007 – Q4 2015; marked 'Post' in the table) the introduction of Family Start. Outcomes are presented for the period covering the child's first, second, and sixth year of life. For each group, the table shows the percentage point difference between the pre- and post-periods in the proportion of individuals with the recorded outcome (labelled 'Difference'). Finally, the rightmost column shows these differences for the treatment group relative to the difference recorded for the control group (i.e., the estimate of difference in difference; or DiD).[56] Note that this table presents descriptive (raw) differences, and do not control for confounding factors.

First and second year outcomes

In the first two years, outcomes such as Oranga Tamariki notifications (general or family violence-related) and assessments (i.e. referral to the Core Assessment phase), and mental health service use were much more common following the introduction of Family Start in both groups. This could be the result of better data collection practices over time, and/or changes in practices among Family Start providers or in CYF/Oranga Tamariki. Mental health service use is a composite outcome variable, with some underlying variables being available only in later time periods.[57]

In the first year, growth in all Oranga Tamariki-related outcomes grew relatively faster for the treatment group, whereas growth in PHO enrolments and maltreatment-related injuries was relatively slow. With respect to mortality outcomes, reductions in post-neonatal mortality were relatively stronger for the treatment group.[58]

In the second year, the prevalence of Oranga Tamariki-related outcomes was lower overall post introduction compared with the first year outcomes (in both groups). In addition, change in the Oranga Tamariki-related outcomes of Reports of Concern and care placements were lower in the treatment group, leading to a negative DiD estimate. Similarly, a lower increase in the proportion of PHO enrolments and mental health service use was recorded for the treatment group. On the other hand, there was no notable change in maltreatment-related injury hospitalisations; the number of children recording mortality outcomes was too small to be presented and compared.

Sixth year outcomes

In contrast with the first- and second-year outcomes, the Oranga Tamariki-related variables, mental health service use, and maltreatment-related hospitalisation in the sixth year were recorded for a smaller proportion of individuals following the expansion of Family Start (Post). Comparing the two periods, treatment TLAs recorded relatively stronger reductions in maltreatment-related injury hospitalisations, and relatively weaker reductions in maternal mental health service use. Similarly, while PHO enrolments fell in both groups between the two periods, the reduction in treatment TLAs was weaker.

Finally, reductions in overall post-neonatal mortality (between the third and sixth years of life) were recorded in both control and treatment TLAs when comparing the pre and post periods. However, the reduction for treatment TLAs was relatively lower, resulting in a positive DiD estimate.

---

[56] Note that due to some events being very rare, some proportions and differences could not be presented as they compromised protection of child identify.

[57] For example, interactions with mental health services captured in PrimHD are collected from 2007, whereas mental health hospitalisations cover the entire period.

[58] Post-neonatal Injury mortality outcome is not presented, in accordance with Statistic New Zealand's confidentiality protocols as the numbers were below the allowed reporting threshold.

Overall, the lower rates of many outcomes in the period prior to the introduction of Family Start raises a concern regarding the ability of DiD to identify the effects of the programme. While these lower rates could simply reflect less common interaction of children and mothers with these services, it could also signal compromised data quality, especially in the pre-treatment period. If the latter possibility is the case, and the effect is the same across the two groups, then our estimates may be reduced in magnitude and estimated less precisely. [59] On the other hand, if this error affects one group differently than the other, then the estimates may result in an upward or downward artificial bias.

[59] Note that if there is a data quality issue during a specific period, then some of the error introduced will be controlled for by the birth cohort fixed effects. Also, if there is a data quality issue that impacts some TLAs which is consistent over time then this issue will be controlled for by birth-TLA fixed effects.

Table 5. Proportion of treatment and control groups with recorded outcomes, Q1 2003 - Q2 2005 (Pre) and Q3 2007 - Q4 2015 (Post)

	Control			Treatment			
First year	Pre	Post	Difference	Pre	Post	Difference	*DiD*
Number of observations	5,874	25,953	-	8,475	36,147	-	-
Post neonatal mortality	0.0031	0.0030	-0.0001	0.0046	0.0035	-0.0011	-0.0011
Post neonatal injury death	-	0.0010	-	0.0007	0.0007	0.0000	-
Post neonatal SUDI	0.0020	0.0016	-0.0004	0.0025	0.0017	-0.0007	-0.0003
OT/CYF Report of Concern or other initial intake event	0.1619	0.2733	0.1114	0.1352	0.2532	0.1180	0.0066
OT/CYF assessment	0.1006	0.1906	0.0900	0.0846	0.1821	0.0975	0.0075
OT/CYF care placement	0.0209	0.0229	0.0019	0.0181	0.0206	0.0025	0.0006
Child or sibling appeared in a Police FV report to OT/CYF	0.0460	0.2074	0.1614	0.0404	0.2204	0.1800	0.0186
Hospitalised for a maltreatment related injury	0.0020	0.0035	0.0014	0.0039	0.0046	0.0008	-0.0007
Hospitalised for a long bone fracture	-	0.0012	-	0.0014	0.0022	0.0007	-
Enrolled with a PHO	0.9372	0.9711	0.0339	0.9391	0.9544	0.0152	-0.0187
Mother received publicly funded mental health services	0.0684	0.2491	0.1807	0.0411	0.1582	0.1171	-0.0635
Second year	Pre	Post	Difference	Pre	Post	Difference	*DiD*
Post neonatal mortality	0.0010	0.0006	-0.0004	-	0.0007	-	-
Post neonatal injury death	-	-	-	-	0.0003	-	-
OT/CYF Report of Concern or other initial intake event	0.1542	0.2180	0.0638	0.1384	0.1971	0.0587	-0.0051
OT/CYF assessment	0.1134	0.1369	0.0235	0.1062	0.1312	0.0250	0.0015
OT/CYF care placement	0.0204	0.0184	-0.0020	0.0198	0.0164	-0.0034	-0.0013
Child or sibling appeared in a Police FV report to OT/CYF	0.0689	0.2212	0.1523	0.0701	0.2208	0.1508	-0.0015
Hospitalised for a maltreatment related injury	0.0041	0.0055	0.0015	0.0046	0.0057	0.0011	-0.0003
Enrolled with a PHO	0.9668	0.9829	0.0161	0.9618	0.9714	0.0097	-0.0064
Mother received publicly funded mental health services	0.1113	0.2458	0.1344	0.0605	0.1501	0.0896	-0.0448
Sixth year	Pre	Post	Difference	Pre	Post	Difference	*DiD*
Post neonatal mortality (third to sixth years)	0.0015	0.0002	-0.0013	0.0014	0.0003	-0.0011	0.0002
Post neonatal injury death (third to sixth years)	-	-	-	0.0011	-	-	-
OT/CYF Report of Concern or other initial intake event	0.1992	0.1336	-0.0656	0.2007	0.1146	-0.0861	-0.0205
OT/CYF assessment	0.1103	0.0736	-0.0367	0.1165	0.0695	-0.0470	-0.0103
OT/CYF care placement	0.0158	0.0116	-0.0043	0.0156	0.0095	-0.0061	-0.0018
Child or sibling appeared in a Police FV report to OT/CYF	0.1537	0.1483	-0.0054	0.1802	0.1497	-0.0305	-0.0250

Hospitalised for a maltreatment related injury	0.0082	0.0055	-0.0026	0.0131	0.0072	-0.0059	-0.0033
Enrolled with a PHO	0.9852	0.7689	-0.2163	0.9837	0.7731	-0.2106	0.0056
Mother received publicly funded mental health services	0.1992	0.1726	-0.0266	0.1235	0.1056	-0.0180	0.0086

Notes: Percentages derived from fewer than 6 observations have been supressed. For full list of variables, see Appendix C.

Strengths and limitations

Key strengths of the evaluation approach and methodology include the following:

- He Awa Whiria. Using He Awa Whiria in the evaluation has enabled the impact of Family Start to be assessed using Māori, Pasifika and Pākehā frameworks. This allowed for interpretation of evaluation findings through Te Ao Māori and Pasifika worldviews, rather than using an exclusively Pākehā framework.
- Relatively large sample sizes. Due to the use of the IDI, the sample sizes used in the current evaluation were larger than is typical for programme evaluations in the social services. This increases the statistical power to detect a significant treatment effect (although this is dampened for rare outcomes such as mortality, as outlined in the limitations section below).
- Longitudinal data. The outcomes data sourced from the IDI were longitudinal, allowing for a more robust analysis of the maintenance of treatment effects over time. This also allowed for tighter controls around the direction of causality, which is an issue with cross-sectional research designs.
- Access to administrative data from a range of agencies. This is another benefit of the IDI.

Limitations of the evaluation approach include the following:

- No randomised assignment. The methodological approach for this study did not involve randomised assignment to treatment and control groups. This limits the ability to draw robust causal inferences about the impact that Family Start has on enrolled children and their whānau.
- Deficit focus. The data contained in the IDI are largely drawn from administrative data related to service provision and interventions carried out by government agencies. The identified outcome measures therefore do not directly capture indicators of wellbeing, particularly as conceptualised using a Māori or Pasifika knowledge framework.
- Identifying ethnicity. Although a 'total ethnicity' approach was used in the analysis, we acknowledge the inherent limitations of administrative ethnicity data, in that not all individuals will be categorised in a way that aligns with their own understanding of their ethnicity.
- Reliance on service engagement. We were not able to use survey data in the analysis (including survey data held in the IDI), as the number of respondents from the Family Start population was too low to provide robust data. This means that useful measures such as subjective wellbeing (captured by the General Social Survey) could not be incorporated into analysis.
- Safeguarding effects. Some of the selected outcome measures (e.g., Oranga Tamariki notifications, mental health service use) may be subject to a 'safeguarding effect', whereby engagement with Family Start increases identification of, and service referral for, particular whānau issues that would otherwise have gone unaddressed. This could have the effect of increasing the rate of outcomes that may be perceived as negative for children and whānau and hence care must be taken in interpreting results.
- Bias in service engagement data for Māori. Previous research has indicated that Māori, and to a lesser extent Pasifika, individuals and whānau may be more likely to be subject to the attention of government agencies (and therefore appear in their datasets) than non-Māori individuals and families.[60] Given that many outcomes measured by government agencies and

[60] Milne, B., Li, E., & Sporle, A. (2020). *Intergenerational analyses using the IDI: An update*. COMPASS Research Centre.

included in the IDI are deficit-framed, this may skew results in a more negative direction for Māori compared with other ethnicities (potentially via a more intense safeguarding effect).

- Power to detect differences in outcomes. Some of the outcomes measured in this study were relatively rare events, such as measures of maltreatment and mortality. Because of the low base-rates of these outcomes in the general population, the statistical power to detect any effect of Family Start on these outcomes was low. This means that care needs to be taken when interpreting results relating to these outcomes, because lack of statistical significance may not indicate an absence of an effect.
- Measuring a wide array of effects. Care must be taken when interpreting the size of the effects estimated in the current study. This is because it is unlikely that all whānau who engaged with Family Start required support in all of the areas that were able to be addressed by the programme, but instead required more specific support in some wellbeing domains than others. This can cause overall effect sizes to be smaller for each outcome, because the effect is averaged across all families who engaged with Family Start, not just families who required or received support in each domain.
- Limited timeframes of available data. In some cases, administrative datasets have only been added to the IDI relatively recently, which means that outcomes are only available for children born more recently (for example ECE and B4SC data).
- Lack of information about engagement and exact interventions received. The Family Start data available in the IDI do not contain information on the exact interventions received by children and whānau or the level of engagement. Effects were therefore only able to be estimated at the programme level, rather than the intervention level. We explored whether there was value in stratifying analyses by the length of time a child participated in the programme but concluded that this was a poor indicator of engagement.
- Delivery of other services to participants. It is likely that individuals in both the treatment and control groups received assistance through services outside of Family Start. However, information about these other services was not available in the IDI.
- Presence of Family Start participants in control samples. Additionally, due to the use of propensity to participate as the sole proxy for programme eligibility in the DiD analyses, some individuals who participated in Family Start are included in control samples. This is expected to reduce the magnitude of the estimated treatment effect.
- Multiple comparisons. The current evaluation explores the impact of Family Start on numerous outcomes, which increases the risk of false positives (i.e. Type I error). We have not adopted one of the formal methods available for explicitly addressing this issue and simply note it here.
- Lack of clarity about policy-significant effect size. We were unable to identify information or documentation that provides clarity on what a meaningful effect size might be for many of the measures of interest. This provided a challenge when attempting to make evaluative judgements about the success of Family Start.

RESULTS

Overview

This section presents separate results from the PSM and DiD analyses for the three groups of outcomes. Results tables present outcomes estimated at the level of the individual child, with the PSM and DiD analyses measuring the impact of Family Start using slightly different lenses.

- Outcomes measured by the PSM analyses reflect the estimated impact for children participating in Family Start. This covers participants born between 2009 and 2015 in TLAs where the programme was officially available.
- Outcomes measured by the DiD analyses reflect the estimated impact of the availability of Family Start on all target children born between 2003 and 2015 (i.e. regardless of whether they participated in Family Start). Note that the focus of the analysis was on the 14 TLAs where Family Start became available between 2005 and 2007.

As discussed above, an important caveat when comparing the PSM and DiD estimates is that in contrast to PSM estimates being more direct treatment effects estimates, the DiD approach provides estimates of 'intention to treat' effects. Given low participation rates in the DiD target population, all else being equal, the latter are expected to be smaller.

The results found below present mortality outcomes per 1,000 children, and other outcomes per 100 children. We adopted this approach from Vaithianathan et al. (2016) as this approach produces results that are more easily discussed (particularly in an evaluative context). In addition, we also provide estimates of the relative impact in terms of a percentage change.[61] Table D6 (Appendix D) provides information on participants' mean outcomes used to calculate the relative impacts for the PSM analyses, and Table E5 (Appendix E) presents the means used to calculate the relative impacts for the DiD analyses.[62]

Confidence intervals and statistical significance

Note that confidence intervals reported in the results tables below are standard 95% confidence intervals (CIs). All impacts that were statistically significant at the p<.05 level are identified in the text below. We also note instances where our results were significant at the p<.10 level (or 90% confidence level) when making comparisons with results from the previous study that were statistically significant at the p<.05 level. Otherwise, results that were only significant at the p<.10 level are not commented on. This is because p<.05 (or 95% confidence level) is the commonly accepted cut-off for statistical significance.

[61] In the PSM analyses, the estimated percentage impact is calculated relative to the control group's mean. In the DiD analyses, the percentage impact is calculated relative to the overall mean.

[62] Only for the outcomes that were found to be statistically significant in the main analysis.

Post-Neonatal Mortality

Post-neonatal mortality: Any ethnicity

Table 6 provides PSM- and DiD-based estimates of the impact of Family Start for the whole sample (i.e. children of any ethnicity) on overall **post-neonatal mortality**, **post-neonatal SUDI** (Sudden Unexplained Death in Infancy), and **post-neonatal injury related deaths**.

The PSM analysis found statistically significant reductions in all three mortality outcomes in the first year of life. The estimates suggested a reduction in overall **post-neonatal mortality** among participants by 42%, or 1.2 per 1,000 children (reduced from 2.9 to 1.7 per 1,000 children), a reduction in **post-neonatal SUDI** by 51%, or 0.7 per 1,000 children (reduced from 1.3 to 0.6 per 1,000 children), and a reduction in **post-neonatal injury-related deaths** among participants by 67%, or 0.6 per 1,000 children (reduced from 0.8 to 0.3 per 1,000 children). Family Start was not found to significantly reduce post-neonatal mortality in the second year of life, nor the third to sixth years (24 to 72 months[63]). These results are broadly consistent with those reported in Vaithianathan et al. (2016).[64]

The DiD analyses did not find evidence that the introduction of Family Start significantly reduced post-neonatal mortality. Although all first-year estimates were negative, indicating a reduction in mortality, there were wide confidence intervals around each estimate indicating a low level of reliability. This is not consistent with Vaithianathan et al. (2016), who reported a statistically significant reduction of 3.5 per 1,000 children (95%CI, 0.3 to 6.7) in post-neonatal mortality in year one. Here, our estimated reduction was smaller (0.8 per 1,000 children, 95%CI, -1.4 to 3.0) and not statistically significant. Note that we used a different study period, sample, and model specification.[65]

Table 6. Impact of Family Start on post-neonatal mortality (Any ethnicity)

	Outcome period		
	First year	Second year	Third to sixth year
Propensity Score Matching estimates			
Post neonatal mortality	**-0.00120****	-0.00035	0.00024
	[-0.00212, -0.00029]	[-0.00083, 0.00013]	[-0.00086, 0.00133]
Post neonatal injury death	**-0.00056****	-0.00002	0.00031
	[-0.00100, -0.00013]	[-0.00033, 0.00030]	[-0.00033, 0.00096]
Post neonatal SUDI	**-0.00066****		
	[-0.00125, -0.00005]		
Difference in Difference estimates			
Post neonatal mortality	-0.00080	0.00061	0.00011
	[-0.00304, 0.00144]	[-0.00023, 0.00146]	[-0.00111, 0.00134]

[63] Mortality outcomes were measured over years three to six rather than just for year six (which was the case for other outcomes). This approach was adopted to give mortality analyses greater statistical power due to the rareness of childhood mortality for children in this age range.

[64] Note that Vaithianathan et al. (2016) found a small reduction in the second year. This was significant at the 10% level, and was not replicated in this analysis.

[65] However, when we re-estimated the model with a start period of 2004q3 as in Vaithianathan et al. (2016) (rather than 2003q1) we found that overall post-neonatal mortality and post-neonatal injury-related deaths were statistically significant, and both pass the parallel trends tests. Please see Appendix E for further discussion.

Post neonatal injury death	-0.00067*	0.00039*	0.00009
	[-0.00145, 0.00011]	[-0.00006, 0.00083]	[-0.00063, 0.00081]
Post neonatal SUDI	-0.00036		
	[-0.00209, 0.00137]		

Note: Parameter estimates statistically different from zero at 99% (***), 95% (**), 90% (*) confidence.

Post-neonatal mortality: Māori

Table 7 presents the PSM- and DiD-results for Māori children. The PSM analysis suggests that in the first year of life, Family Start reduced **post-neonatal SUDI** among Māori participants by 59%, or 1.0 per 1,000 children (reduced from 1.7 to 0.7 per 1,000 children) and reduced **post-neonatal injury-related deaths** among Māori participants by 63%, or 0.8 per 1,000 children (reduced from 1.3 to 0.5 per 1,000 children).[66]

Family Start was not found to significantly reduce post-neonatal mortality for Māori participants in the second year of life, nor between the third and sixth years of life. These results are similar to those found in Vaithianathan et al. (2016), although the previous study found larger and more significant impacts in the first year for Māori children.[67]

The DiD results did not find any statistically significant reductions in any of the mortality outcomes.[68] This contrasts with Vaithianathan et al. (2016), who found a statistically significant reduction in overall **post-neonatal mortality** among Māori children in the first year of life and reported a larger estimate (reduction of 4.9 per 1,000 children, compared with 0.1 per 1,000 children in this analysis).

Table 7. Impact of Family Start on post-neonatal mortality (Māori)

	Outcome period		
	First year	Second year	Third to sixth year
Propensity Score Matching estimates			
Post neonatal mortality	-0.00129*	-0.00024	0.00008
	[-0.00266, 0.00008]	[-0.00102, 0.00053]	[-0.00143, 0.00159]
Post neonatal injury death	**-0.00080****	0.00008	0.00060
	[-0.00155, -0.00006]	[-0.00044, 0.00060]	[-0.00030, 0.00149]
Post neonatal SUDI	**-0.00103****		
	[-0.00192, -0.00005]		
Difference in Difference estimates			
Post neonatal mortality	-0.00193	0.00072	0.00006
	[-0.00568, 0.00183]	[-0.00032, 0.00176]	[-0.00159, 0.00170]
Post neonatal injury death	-0.00109	0.00036	0.00026
	[-0.00246, 0.00027]	[-0.00028, 0.00099]	[-0.00084, 0.00137]
Post neonatal SUDI	-0.00110		

66 The PSM analysis also found a reduction in overall post-neonatal mortality among Māori participants by 37%, or 1.3 per 1,000 children (reduced from 3.5 to 2.2 per 1,000 children) in the first year of life. However, this was significant at the p<.10 or 90% confidence level.

67 Vaithianathan et al. (2016) also found significant impacts at the 10% level in the second year that were not replicated in the current study.

68 Using the alternative 'benefit' based target group, our analysis estimates a (non-significant) reduction in year one overall post-neonatal mortality of 2.2 per 1,000 children.

[-0.00415, 0.00195]

Note: Parameter estimates statistically different from zero at 99% (***), 95% (**), 90% (*) confidence.

Post-neonatal mortality: Pasifika

Table 8 presents the PSM- and DiD-results on mortality outcomes for Pasifika children. In the PSM analyses, Family Start was found to significantly reduce overall **post-neonatal mortality** among Pasifika participants by 62%, or 2.0 per 1,000 children (reduced from 3.1 to 1.2 children per 1,000) in the first year of life.

There were no statistically significant reductions found in **SUDI** or **injury-related deaths** among Pasifika participants in the first year of life. There were also no significant reductions in any mortality-related outcomes in the second year of life, or between the third and sixth years for Pasifika participants.

The significant reduction in post-neonatal mortality in the first year of life replicates a similar finding in Vaithianathan et al. (2016), who found a reduction of 2.3 per 1,000 children. However, the significant reductions in **post-neonatal SUDI** (1.4 per 1,000 children) and **post-neonatal injury-related deaths** (1.1 per 1,000 children) found in the first year in the previous study were not replicated in the current study.

The DiD analyses did not find that the introduction of Family Start significantly reduced post-neonatal mortality for Pasifika children.[69] This is consistent with the findings of Vaithianathan et al. (2016), who also did not find significant reductions in post-neonatal mortality for Pasifika children in their DiD analyses.

Table 8. Impact of Family Start on post neo-natal mortality (Pasifika)

	Outcome period		
	First year	Second year	Third to sixth year
Propensity Score Matching estimates			
Post neonatal mortality	**-0.00195*** **	-0.00011	0.00013
	[-0.00341, -0.00050]	[-0.00069, 0.00047]	[-0.00163, 0.00189]
Post neonatal injury death	-0.00038	-0.00017*	0.00000
	[-0.00100, 0.00025]	[-0.00036, 0.00002]	[-0.00017, 0.00017]
Post neonatal SUDI	-0.00061		
	[-0.00161, -0.00005]		
Difference in Difference estimates			
Post neonatal mortality	0.00366*	0.00033	-0.00013
	[-0.00042, 0.00774]	[-0.00154, 0.00220]	[-0.00235, 0.00209]
Post neonatal injury death	0.00119	0.00092	0.00099
	[-0.00108, 0.00346]	[-0.00074, 0.00259]	[-0.00070, 0.00268]
Post neonatal SUDI	0.00290		
	[-0.00085, 0.00665]		

Note: Parameter estimates statistically different from zero at 99% (***), 95% (**), 90% (*) confidence.

[69] Estimates for the 'benefit' based target group were very similar: see table for Table E6-

Table E8 for more information.

Post-neonatal mortality: summary

In summary, our PSM analyses found that Family Start significantly reduced participants' overall post-neonatal mortality, post-neonatal mortality due to injury, and post-neonatal SUDI in the first year of life. Significant reductions in two mortality outcomes were also observed for Māori participants in the first year of life (post-neonatal mortality due to injury, and post-neonatal SUDI). For Pasifika participants, there was a significant reduction in overall post-neonatal mortality in the first year of life. On the other hand, while the year one DiD analyses indicated a reduction in mortality outcomes, none of these estimates were statistically significant.

As per the rubric presented in Appendix A for evaluating the impact of Family Start, the overall evidence suggests that Family Start is 'meeting expectations' with respect to mortality-related impacts. This evaluative judgement is based on the assumption that the PSM approach is valid in the present study (i.e. no selection bias), and that the lack of findings in the DiD reflects the practical limitations of applying the approach this context.

Note that this judgement is made on the strength of evidence and the precision of the estimates (i.e. the size of the confidence intervals), and is not a reflection on the substantial importance of finding that Family Start has a measurable impact on reducing the number of infant deaths for its participants.

Health and education-related outcomes

Health and education-related outcomes: Any ethnicity

Table 9 presents the PSM- and DiD-estimates of the impact of Family Start on engagement with health services and participation in Early Childhood Education (ECE) for the whole sample (i.e. any ethnicity).

Results from the PSM analyses indicated significant increases in Family Start participants' likelihood of being enrolled with a PHO, being fully immunised at each milestone age, and attending a Before School Check (B4SC). These are in line with the programme's short-term goals of enrolling children with a PHO, and ensuring immunisations are up-to-date.

Family Start was found to increase the likelihood of participants' PHO enrolment by the end of the first year by 1.7%, or 1.6 per 100 children (increased from 95.9 to 97.5 per 100 children); and an increase of 0.7%, or 0.70 per 100 children (increased from 98.3 to 99.0 per 100 children) by the end of the second year. On the other hand, there was no significant difference in PHO enrolment by the end of the sixth year.

With respect to immunisations, the PSM analyses found that the programme increased participants' rates of full immunisation at one or more milestone ages in the first year by 4.6%, or 4.0 per 100 children (increased from 87.8 to 91.8 per 100 children); and in the second year by 3.3%, or 3.0 per 100 children (increased from 89.4 to 92.4 per 100 children). Furthermore, the PSM analyses found that the programme increased participants' rates of full immunisation at every milestone age in the first year by 6.1%, or 3.7 per 100 children (increased from 60.6 to 64.3 per 100 children); by 4.7%, or 3.5 per 100 children (increased from 74.7 to 78.3 per 100 children) in the second year; and by 3.8%, or 3.1 per 100 children (increased from 79.0 to 82.1 per 100 children) by the sixth year.[70]

The PSM analysis also found that the programme increased the likelihood of attending the B4SC by 2.8%, or 2.2 per 100 children (increased from 78.5 to 80.7 per 100 children). Limiting analysis to B4SC attendees, Family Start participants were 3.5 and 4.3 per 100 children more likely to record one or more significant and non-significant issues, respectively. It is plausible that the success of Family Start in increasing enrolment has resulted in the participation of families that face greater challenges and thus we observe a greater proportion of children with significant and non-significant issues. This could lead to biased estimates since PSM matches children's characteristics at (and before) the time of birth, and not at the time the assessment was made.

While recording a positive estimate, the PSM analysis did not find any statistically significant differences in the likelihood of children having ever enrolled with an ECE provider by age 6. During the study period some Family Start participants received financial assistance towards ECE costs (Early Learning Payment). The Early Learning Payment is a subsidy provided to families enrolled with FS that could be used to defray the cost of ECE for children between 18 months and 3 years. This scheme was only made available in some of the TLAs in the study period but is now available across all Family Start sites. Measures of the timing and duration of ECE participation were not examined, and as a result we are unable to say whether or not enrolment was increased at ages when the Early Learning Payment is available, or whether the duration of ECE participation was increased. The PSM ECE result also contrasts with Vaithianathan et al. (2016), who found a much larger and statistically significant increase in ECE enrolments as measured at the B4SC (7.6 per 100 children compared to 1.1 per 100 children). However, the two studies used different data sources, hence the two outcome measures are not directly comparable.[71]

Finally, the PSM analysis found that participants' mothers' likelihood of receiving publicly-funded mental health and addiction services increased by 50.0%, or 8.7 per 100 children (from 17.4 to 26.1 per 100 children) in the first year of the child's life; by 59.5%, or 6.9 per 100 children (from 17.4 to 24.2 per 100 children) in the second year of the child's life; and by 23.0%, or 3.4 per 100 children (from 14.9 to 18.3 per 100 children) in the child's sixth year of life.

Note that these results are not necessarily negative, as these estimates could also reflect a safeguarding effect. That is, issues such as maternal mental health conditions are more likely to be referred to agencies - and therefore appear in official records - when mothers are involved in Family Start (e.g., Family Start may be ensuring that mothers receive support for mental health and addiction where it might otherwise have gone undetected). This explanation is also consistent with Family Start's intended short-term outcome relating to the identification of, and support for, child and parent physical and mental health issues. The possibility of a safeguarding effect was also discussed in Vaithianathan et al. (2016), who also found increased likelihood for similar outcomes.[72] Following a supplementary analysis, they found that in the first year, mothers of children enrolled pre-birth were more likely to use mental health services, but not addiction

[70] There is one milestone age at 60 months.

[71] We used ECE information from the Ministry of Education, while Vaithianathan et al. (2016) measured ECE enrolment using Ministry of Health data, and only for children that attended the B4SC.

[72] Vaithianathan et al. (2016) examined the impacts on the use of mental health services and use of addiction services separately. They found a significant increase in participants' mothers' use of mental health services of 14.6 per 100 children in the first year, and 8.0 per 100 children in the second year. They also found a significant increase in the use of addiction services of 2.5 per 100 children in first year and in the second year.

services. This provides some support for the suggestion that a safeguarding effect may be influencing the results, though we cannot conclusively determine the extent of this effect in the present report.

The DiD analyses found statistically significant *reductions* in the likelihood of participants' mothers' receiving publicly-funded **mental health and addiction services** in the child's first and second year. The estimates suggested a decrease of 14.2%, or 2.3 per 100 children, in the first year of the child's life, and a decrease of 7.6%, or 1.2 per 100 children, in the second year of the child's life. No statistically significant difference was observed in the sixth year of the child's life. However, parallel trends tests indicated that these outcomes are more likely a continuation of pre-treatment trends, rather than outcomes that can be attributed to the programme.[73] This is in line with the findings of Vaithianathan et al. (2016), who found no significant differences in mental health or addiction service use among mothers in their DiD analyses. Finally, the DiD analyses did not find any significant differences in PHO enrolments among the target group, which also replicated the non-significant findings in Vaithianathan et al. (2016).

Note that DiD analyses did not include B4SC or immunisation outcomes, nor ECE enrolment rates due to data availability limitations in the IDI.

Table 9. Impact of Family Start on health and education-related outcomes (Any ethnicity)

	Outcome period		
	First year	Second year	Sixth year
Propensity Score Matching estimates			
Enrolled with a PHO	**0.01619*****	**0.00701*****	0.00335*
	[0.01171, 0.02066]	[0.00321, 0.01081]	[-0.00053, 0.00723]
Fully immunised at 1+ milestone age	**0.04002*****	**0.02962*****	
	[0.03375, 0.04628]	[0.02352, 0.03572]	
Fully immunised at every milestone age	**0.03669*****	**0.03536*****	**0.03075*****
	[0.02746, 0.04592]	[0.02692, 0.04379]	[0.02099, 0.04050]
Enrolled with an ECE provider			0.01053*
			[-0.00053, 0.02158]
Attended B4SC			**0.02189*****
			[0.01251, 0.03127]
B4SC identified at least one significant issue			**0.03464*****
			[0.02334, 0.04593]
B4SC identified at least one non-significant issue			**0.04326*****
			[0.03229, 0.05424]
B4SC indicated a healthy BMI			-0.00018
			[-0.00821, 0.00784]
Mother received publicly funded mental health services	**0.08702*****	**0.06853*****	**0.03423*****
	[0.07919, 0.09484]	[0.06076, 0.07629]	[0.02589, 0.04257]
Difference in Difference estimates			
Enrolled with a PHO	-0.01165	-0.00311	-0.00141
	[-0.03098, 0.00769]	[-0.01261, 0.00638]	[-0.00408, 0.00126]
Mother received publicly funded mental health services	**-0.02251*****	**-0.01247****	0.00532
	[-0.03571, -0.00930]	[-0.02255, -0.00239]	[-0.00530, 0.01594]

Note: Parameter estimates statistically different from zero at 99% (***), 95% (**), 90% (*) confidence.

[73] See Table E5 for more information.

Health and education-related outcomes: Māori

Table 10 presents the PSM and DiD-estimated impacts of Family Start on engagement with health services and participation in Early Childhood Education (ECE) for Māori children.

The PSM analyses indicated a statistically significant increase in Māori participants' **PHO enrolment** rates in the first year by 2.0%, or 1.9 per 100 children (from 95.7 to 97.6 per 100 children); and by 0.8%, or 0.8 per 100 children in the second year (from 98.3 to 99.1 per 100 children). However, there was no significant difference in PHO enrolment in the sixth year.

The PSM analyses also found a statistically significant increase in Māori participants' rates of **full immunisation at one or more milestone ages** in the first year by 5.1%, or 4.4 per 100 children (from 85.6 to 90.0 per 100 children); and by 3.5%, or 3.1 per 100 children in the second year (from 88.4 to 91.5 per 100 children). Similarly, a statistically significant increase was also recorded for Māori participants' rates of **full immunisation at every milestone age** by 7.7%, or 4.1 per 100 children (from 53.2 to 57.3 per 100 children) in the first year; by 5.8%, or 4.1 per 100 children (from 70.4 to 74.5 per 100 children) in the second year; and by 4.6%, or 3.6 per 100 children in the sixth year (from 77.6 to 81.2 per 100 children).

With respect to **B4SC attendance**, the PSM analysis found that the programme increased Māori participants' likelihood of attendance by 3.1%, or 2.4 per 100 children (from 76.8 to 79.2 per 100 children). This is not consistent with Vaithianathan et al. (2016), which did not find any significant increase. No significant difference in the rate of **ECE enrolment** was found for Māori participants.

Consistent with the findings for the total sample, the PSM analyses found that Family Start participation increased the likelihood that Māori participants who completed a B4SC recorded **significant** and **non-significant issues**,[74] and also increased the likelihood of that participants' mothers received publicly-funded **mental and addiction services**.[75]

Finally, no significant differences in health or education-related outcomes were identified for Māori target groups in the DiD analyses. This is in line with the findings in Vaithianathan et al. (2016).

Table 10. Impact of Family Start on health and education-related outcomes (Māori)

	Outcome period		
	First year	Second year	Sixth year
Propensity Score Matching estimates			
Enrolled with a PHO	**0.01885***** [0.01301, 0.02470]	**0.00792***** [0.00300, 0.01283]	0.00362 [-0.00120, 0.00843]
Fully immunised at 1+ milestone age	**0.04382***** [0.03514, 0.05250]	**0.03061***** [0.02240, 0.03882]	
Fully immunised at every milestone age	**0.04129***** [0.02879, 0.05379]	**0.04144***** [0.02991, 0.05297]	**0.03557***** [0.02262, 0.04853]
Enrolled with an ECE provider			0.00623 [-0.00827, 0.02072]

74 An increase in the likelihood of recording one or more significant issues in the B4SC by 17.1%, or 4.2 per 100 children (from 24.5 to 28.7 per 100 children), and an increase in the likelihood of recording a non-significant issue in the B4SC by 17.2%, or 4.4 per 100 children (from 25.6 to 30.0 per 100 children).

75 An increase in the likelihood that Māori participants' mothers received publicly-funded mental health and addiction services by 44.8%, or 7.7 per 100 children (from 17.2 to 24.9 per 100 children) in the first year of the child's life; by 33.5%, or 5.8 per 100 children (from 17.3 to 23.1 per 100 children) in the second year of the child's life; and by 20.8%, or 3.2 per 100 children (from 15.4 to 18.6 per 100 children) in the child's sixth year of life.

Attended B4SC			**0.02380***
			[0.01138, 0.03623]
B4SC identified at least one significant issue			**0.04199***
			[0.02694, 0.05705]
B4SC identified at least one non-significant issue			**0.04424***
			[0.02987, 0.05861]
B4SC indicated a healthy BMI			-0.00174
			[-0.01248, 0.00901]
Mother received publicly funded mental health services	**0.07706***	**0.05775***	**0.03191***
	[0.06679, 0.08732]	[0.04755, 0.06795]	[0.02077, 0.04305]
Difference in Difference estimates			
Enrolled with a PHO	-0.00735	-0.00185	-0.00138
	[-0.02667, 0.01197]	[-0.01337, 0.00967]	[-0.00491, 0.00214]
Mother received publicly funded mental health services	-0.00977*	-0.00315	0.00630
	[-0.02066, 0.00113]	[-0.01375, 0.00745]	[-0.00607, 0.01867]

Note: Parameter estimates statistically different from zero at 99% (***), 95% (**), 90% (*) confidence.

Health and education-related outcomes: Pasifika

Table 11 presents the PSM and DiD-estimated impacts of Family Start on engagement with health services and participation in Early Childhood Education (ECE) for Pasifika children.

The PSM analyses indicated an increase in the likelihood of Pasifika participants' **PHO enrolment** in the first year by 1.3%, or 1.2 per 100 children (from 95.5 to 96.7 per 100 children); and by 0.8%, or 0.8 per 100 children in the second year (from 97.9 to 98.7 per 100 children). No significant difference in PHO enrolment in the sixth year was recorded.

The PSM analyses also found that Pasifika participants' likelihood of **full immunisation at one or more milestone ages** in the first year increased by 3.7%, or 3.3 per 100 children (from 89.5 to 92.8 per 100 children); and by 2.3%, or 2.1 per 100 children in the second year (from 90.1 to 92.2 per 100 children). Similarly, the analyses suggested an increase in Pasifika participants' likelihood of **full immunisation at every milestone age** in the first year by 5.3%, or 3.3 per 100 children (from 62.0 to 65.3 per 100 children); and by 3.2%, or 2.4 per 100 children in the second year (from 75.0 to 77.4 per 100 children). However, no significant increase in the likelihood of achieving full immunisation at every milestone age in the sixth year was detected.

With respect to B4SC attendance, the PSM analyses found that the programme increased the likelihood of Pasifika participants **attending the B4SC** by 2.8%, or 2.1 per 100 children (from 74.2 to 76.3 per 100 children). On the other hand, there was no significant difference in the rate of **ECE enrolment** for Pasifika participants. Again, as for the total sample and Māori participants, the PSM analyses found that Pasifika participants who completed the B4SC were more likely to record one or more **significant** and **non-significant issues** in the B4SC,[76] and Pasifika participants' mothers were more likely to receive publicly-funded **mental health and addiction services.**[77]

[76] An increase in the likelihood of Pasifika participants having significant issues identified in the B4SC of 12.0%, or 2.6 per 100 children (from 21.6 to 24.2 per 100 children), and of 17.5%, or 3.3 per 100 children (from 18.9 to 22.2 per 100 children) for non-significant issues.

[77] An increase of 69.9%, or 6.5 per 100 children (from 9.3 to 15.8 per 100 children) in the first year of the child's life; and of 52.2%, or 4.8 per 100 children (from 9.2 to 14.0 per 100 children) in the second year of the child's life. There was no significant difference in the child's sixth year of life.

Finally, no significant differences in health or education-related outcomes were identified for the Pasifika target group in the DiD analyses. This is in line with the findings in Vaithianathan et al. (2016).

Table 11. Impact of Family Start on health and education-related outcomes (Pasifika)

	Outcome period		
	First year	Second year	Sixth year
Propensity Score Matching estimates			
Enrolled with a PHO	**0.01195***** [0.00298, 0.02092]	**0.00834**** [0.00052, 0.01615]	0.00460 [-0.00302, 0.01223]
Fully immunised at 1+ milestone age	**0.03322***** [0.02222, 0.04422]	**0.02085***** [0.00973, 0.03197]	
Fully immunised at every milestone age	**0.03278***** [0.01615, 0.04942]	**0.02372***** [0.00828, 0.03917]	0.01574* [-0.00296, 0.03445]
Enrolled with an ECE provider			0.01253 [-0.00792, 0.03298]
Attended B4SC			**0.02058**** [0.00264, 0.03852]
B4SC identified at least one significant issue			**0.02644***** [0.00664, 0.04624]
B4SC identified at least one non-significant issue			**0.03347***** [0.01422, 0.05272]
B4SC indicated a healthy BMI			-0.00818 [-0.02546, 0.00910]
Mother received publicly funded mental health services	**0.06497***** [0.05312, 0.07682]	**0.04756***** [0.03611, 0.05900]	0.00819 [-0.00382, 0.02020]
Difference in Difference estimates			
Enrolled with a PHO	-0.00340 [-0.03399, 0.02719]	0.00564 [-0.01265, 0.02394]	0.00108 [-0.00407, 0.00624]
Mother received publicly funded mental health services	-0.00473 [-0.01871, 0.00926]	-0.00154 [-0.01466, 0.01158]	0.00551 [-0.01497, 0.02599]

Note: Parameter estimates statistically different from zero at 99% (***), 95% (**), 90% (*) confidence.

Health and education-related outcomes: summary

In summary, PSM analyses indicated that Family Start significantly increased participants' likelihood of being **enrolled with a PHO**, being **fully immunised at each milestone age**, and **attending a Before School Check (B4SC)**. These outcomes were also found for the Māori and Pasifika sub-group analyses. These results suggest that Family Start is meeting its short-term outcomes related to 'child's health and safety' outlined in its Theory of Change model, namely that:

- child/family are enrolled with PHO
- child immunisations are up-to-date.

However, the DiD analyses did not find significant differences in the likelihood of PHO enrolment among for target group children (in line with the findings in Vaithianathan et al. [2016]).

Other findings were more difficult to interpret. These included a greater likelihood of Family Start participant children having **significant** and **non-significant issues** identified in the B4SC, and of having mothers who received publicly-funded **mental health and addiction support** (statistically significant in the PSM analyses only). Although these outcomes may appear to indicate a deterioration in circumstances, they could also be interpreted as Family Start meeting targeted short-term outcomes related to 'child's health and safety' as outlined in its Theory of Change model, namely:

- children's health and safety issues are identified early and addressed
- address family violence and alcohol and drug misuse that impact on the child (access to specialist community services)
- identification of and access to services for mothers with post-natal depression.

There was no statistically significant impact on the likelihood of children enrolling with an ECE provider. Overall, measuring against the evaluation rubric presented in Appendix A for evaluating the impact of Family Start, this evidence suggests that Family Start is 'meeting expectations' in terms of some key health and education-related outcomes.

Child protection outcomes

Child protection outcomes: Any ethnicity

Table 12 presents the PSM and DiD-estimated impacts of Family Start on Oranga Tamariki contact, and hospitalisations for maltreatment-related injuries and long bone fractures for the whole sample any ethnicity.

The PSM analyses found that participants in all age groupings were more likely to appear in Oranga Tamariki data than matched control group children. While the magnitude of this impact fell over time, it remained statistically significant at all ages. This included an increase in the likelihood of being a subject of a Report of Concern,[78] assessment,[79] care placement,[80] or a Family Violence related Report of Concern or Contact Record[81] by Oranga Tamariki.

The PSM analyses also found that Family Start participants were more likely to be hospitalised with long-bone fractures[82] and maltreatment related injuries. For maltreatment related injury hospitalisations, this greater likelihood was detected in the first two years, but not the sixth.[83]

[78] The likelihood of being the subject of a Report of Concern increased in the first year by 80.2% (or 16.6 per 100 children, from 20.7 to 37.3 per 100 children), by 60.4% (or 9.9 per 100 children, from 16.4 to 26.3 per 100 children) in the second year, and 32.1% (or 4.3 per 100 children, from 13.4 to 17.7 per 100 children) in the sixth year.

[79] The likelihood of being assessed by Oranga Tamariki in the first year increased by 90.0%, or 13.5 per 100 children (from 15.0 to 28.5 per 100 children); by 67.3%, or 7.2 per 100 children in the second year (from 10.7 to 17.9 per 100 children), and by 32.0%, or 2.4 per 100 children in the sixth year (from 7.5 to 9.9 per 100 children).

[80] The likelihood of being placed into care increased by 37.9% (0.8 per 100 children, from 2.0 to 2.8 per 100 children) in the first year; by 52.6%, or 2.0 per 100 children in the second (from 3.8 to 5.8 per 100 children); and by 53.8%, or 0.6 per 100 children (from 1.2 to 1.8 per 100 children) in the sixth year.

[81] This includes sibling-related events. The likelihood that participants (or their siblings) were recorded in a Family Violence Report of Concern or Contact Record by Police to Oranga Tamariki in the first year increased by 50.3% (9.5 per 100 children, from 18.7 to 28.1 per 100 children); in the second year by 38.9% (7.5 per 100 children, from 19.3 to 26.8 per 100 children); and in the sixth year by 22.6% (3.7 per 100 children, from 16.4 to 20.1 per 100 children).

[82] This was only measured in the first year of life. The analysis suggested an increase in the likelihood that participants were hospitalised for long bone fractures in the first year by 68%, or 1.0 per 1,000 children (from 1.5 to 2.6 per 1,000 children).

[83] In the first year, the likelihood increased by 53% (or by 2.2 per 1,000 children, from 4.1 to 6.3 per 1,000 children); and in the second year by 45% (or 2.1 per 1,000 children, from 4.6 to 6.6 per 1,000 children).

Table 12. Impact of Family Start on child protection outcomes (Any ethnicity)

	Outcome period		
	First year	Second year	Sixth year
Propensity Score Matching estimates			
OT/CYF Report of Concern or other initial intake event	**0.1659*** **	**0.0988*** **	**0.0434*** **
	[0.1572, 0.1747]	[0.0907, 0.1069]	[0.0350, 0.0518]
OT/CYF assessment	**0.1347*** **	**0.0720*** **	**0.0237*** **
	[0.1266, 0.1429]	[0.0650, 0.0791]	[0.01718, 0.03031]
OT/CYF care placement	**0.0077*** **	**0.0203*** **	**0.0063*** **
	[0.0044, 0.0110]	[0.0159, 0.0247]	[0.00335, 0.00920]
Child or sibling appeared in a Police FV report to OT/CYF	**0.0945*** **	**0.0750*** **	**0.0371*** **
	[0.0863, 0.1026]	[0.0669, 0.0832]	[0.02852, 0.04578]
Hospitalised for a maltreatment related injury	**0.00219*** **	**0.00207*** **	-0.00058
	[0.0007, 0.0036]	[0.0006, 0.0036]	[-0.00258, 0.00141]
Hospitalised for a long bone fracture	**0.00103** **		
	[0.00014, 0.00192]		
Difference in Difference estimates			
OT/CYF Report of Concern or other initial intake event	0.01236*	0.00484	-0.01333
	[-0.00034, 0.02506]	[-0.01266, 0.02234]	[-0.02908, 0.00242]
OT/CYF assessment	0.01177*	0.00168	-0.00584
	[-0.00184, 0.02539]	[-0.00861, 0.01196]	[-0.01783, 0.00615]
OT/CYF care placement	0.00261	-0.00008	-0.00121
	[-0.00129, 0.00652]	[-0.00895, 0.00880]	[-0.00475, 0.00233]
Child or sibling appeared in a Police FV report to OT/CYF	**0.01835** **	0.00969	**-0.02081** **
	[0.00343, 0.03328]	[-0.01609, 0.03547]	[-0.03740, -0.00421]
Hospitalised for a maltreatment related injury	-0.00065	0.00115	-0.00186
	[-0.00208, 0.00079]	[-0.00122, 0.00352]	[-0.00626, 0.00255]
Hospitalised for a long bone fracture	0.00082*		
	[-0.00004, 0.00169]		

Notes: Parameter estimates statistically different from zero at 99% (***), 95% (**), 90% (*) confidence.

As with some of the health-related outcomes described in the previous sub-section, these outcomes are difficult to interpret. On the one hand, significant increases in rates of contact with Oranga Tamariki and reports of family violence may indicate deteriorating circumstances. On the other hand, Family Start's Theory of Change includes outcomes related to 'Child's health and safety' where one would potentially expect an increase in Oranga Tamariki contact and family violence reporting if the programme *was* having the intended impact. These include to 'address family violence and alcohol and drug misuse that impact on the child', and 'children's health and safety issues are identified and addressed'.

Viewed through this lens, increased rates of reporting and contact may represent an increased level of willingness or ability of participating families and whānau to report concerns to agencies, and/or for health and safety issues to otherwise come to the attention of agencies. The increase in hospitalisation rates may likewise indicate an increased likelihood of children receiving required treatment, rather than an increase in instances of maltreatment.

Additionally, and as noted by Vaithianathan et al. (2016), PSM estimates of impacts on contact with Oranga Tamariki are likely to reflect reverse causality and cannot be confidently attributed to the programme.[84] This is because contact with Oranga Tamariki may be the reason for referral to the programme. Evidence for this possibility includes the fact that across all Oranga Tamariki-related measures, the estimated increased likelihood is strongest in year one (where most referrals to Family Start occur), and either diminish in the sixth year, or are no longer statistically significant.

Vaithianathan et al. (2016) investigated this possibility by undertaking supplementary analyses restricted to children enrolled in Family Start before birth (about one-fifth of all participants). They found a smaller (but still significant) effect when analysing the likelihood that a participant would be the subject of a Report of Concern, but the impacts on care placements were no longer statistically significant.

It would be possible to address reverse causality by matching children based on characteristics at the time they enrolled in Family Start (and characteristics at the equivalent ages for non-participants). While this would control for the child's (and their siblings') contact with Oranga Tamariki prior to enrolling in Family Start, interpretation would remain challenging as we would still not be able to determine the extent to which the results reflect safeguarding/surveillance effects, and/or changes in whānau/child circumstances.

The DiD analyses, which are not subject to reverse causality, did not find any effect of Family Start on children's likelihood of being the subject of a Report of Concern, being assessed, or placed into care by Oranga Tamariki. There were also no significant differences in the likelihood of hospitalisation for a maltreatment-related injury or a long bone fracture.

However, the DiD analyses found an 11% increase (or 1.8 per 100 children) in the likelihood of children (or their siblings) being recorded in a Family Violence notification by Police to Oranga Tamariki in the first year of life.[85] Conversely, a significant reduction in notifications by Police to Oranga Tamariki (2.1 per 100 children) was identified in the sixth year of life. However, of these two measures, only the first year result satisfied the parallel trends test, indicating that only this finding could be attributed to the programme.

Across both the PSM and DiD analyses, these findings were consistent with those obtained by Vaithianathan et al. (2016) for Oranga Tamariki contact and hospitalisations for maltreatment related injuries and long bone fractures. The previous study did not consider Family Violence notification by Police to Oranga Tamariki.

Child protection outcomes: Māori

Table 13 presents the PSM and DiD-estimated impacts for Māori children. The PSM analyses showed significant increases in the likelihood of Oranga Tamariki contact for Māori participants, but no significant differences in hospitalisations for maltreatment-related injuries or long bone fractures.

---

[84] Note that while reverse causality is most likely to affect year one results, it could potentially (indirectly) affect outcomes in later years. For example, where children had such interactions recorded at an early age (e.g., first year) they are more likely to be monitored over time, and hence reported in later years.

[85] See Table E5 for more information.

More specifically, the PSM analyses suggested that Family Start increased the likelihood that Māori participants were the **subject of a Report of Concern**,[86] **assessment**,[87] or a **care placement**.[88] In addition, participants (and/or their siblings) were more likely to be included in a **Family Violence notification**.[89] As with the overall population, the relative size of this impact decreased over time but remained significant.

Table 13. Impact of Family Start on child protection outcomes (Māori)

	Outcome period		
	First year	Second year	Sixth year
Propensity Score Matching estimates			
OT/CYF Report of Concern or other initial intake event	**0.17581*****	**0.10430*****	**0.04468*****
	[0.16393, 0.18769]	[0.09320, 0.11540]	[0.03327, 0.05610]
OT/CYF assessment	**0.14350*****	**0.07871*****	**0.02222*****
	[0.13243, 0.15457]	[0.06889, 0.08853]	[0.01317, 0.03127]
OT/CYF care placement	**0.00599****	**0.02367*****	**0.00739*****
	[0.00131, 0.01067]	[0.01743, 0.02992]	[0.00319, 0.01158]
Child or sibling appeared in a Police FV report to OT/CYF	**0.09696*****	**0.07830*****	**0.04129*****
	[0.08573, 0.10819]	[0.06697, 0.08963]	[0.02911, 0.05346]
Hospitalised for a maltreatment related injury	0.00203*	0.00215*	-0.00167
	[-0.00012, 0.00418]	[-0.00003, 0.00433]	[-0.00441, 0.00107]
Hospitalised for a long bone fracture	0.00080		
	[-0.00055, 0.00215]		
Difference in Difference estimates			
OT/CYF Report of Concern or other initial intake event	**0.01857****	0.00900	-0.01341
	[0.00199, 0.03516]	[-0.00981, 0.02781]	[-0.03247, 0.00566]
OT/CYF assessment	0.01945*	0.00069	-0.00433
	[0.00031, 0.03859]	[-0.01089, 0.01227]	[-0.01838, 0.00972]
OT/CYF care placement	0.00258	-0.00763	-0.00233
	[-0.00253, 0.00769]	[-0.02034, 0.00508]	[-0.00781, 0.00315]
Child or sibling appeared in a Police FV report to OT/CYF	**0.02322****	0.01918	-0.01952*
	[0.00258, 0.04385]	[-0.01109, 0.04945]	[-0.04172, 0.00269]
Hospitalised for a maltreatment related injury	-0.00086	0.00194	-0.00142
	[-0.00353, 0.00181]	[-0.00157, 0.00545]	[-0.00799, 0.00514]
Hospitalised for a long bone fracture	0.00050		
	[-0.00087, 0.00187]		

Notes: Parameter estimates statistically different from zero at 99% (***), 95% (**), 90% (*) confidence.

[86] The likelihood of being the subject of a Report of Concern increased in the first year by 72.4% (or 17.6 per 100 children, from 24.3 to 41.9 per 100 children); by 53.1% (or 10.4 per 100 children, from 19.6 to 30.0 per 100 children) in the second year; and by 28.5% (or 4.5 per 100 children, from 15.8 to 20.3 per 100 children) in the sixth year.

[87] The likelihood of being assessed by Oranga Tamariki increased in the first year by 8.9% (or 14.4 per 100 children, from 17.8 to 32.2 per 100 children); by 60.3% (or 7.9 per 100 children, from 13.1 to 21.0 per 100 children) in the second year; and by 23.9% (or 2.2 per 100 children, from 9.2 to 11.4 per 100 children) in the sixth year.

[88] The likelihood of being placed into care increased in the first year by 22.2% (or 0.6 per 100 children, from 2.7 to 3.3 per 100 children); by 51.1% (or 2.4 per 100 children, from 4.7 to 7.1 per 100 children) in the second year; and by 46.7% (or 0.7 per 100 children, from 1.5 to 2.2 per 100 children) in the sixth year.

[89] The likelihood that participants (or their siblings) were recorded in a Family Violence Report of Concern or Contact Record by Police to Oranga Tamariki in the first year increased by 41.0% (9.7 per 100 children, from 23.4 to 33.0 per 100 children); in the second year by 32.2% (7.8 per 100 children, from 24.2 to 32.0 per 100 children); and in the sixth year by 19.8% (4.1 per 100 children, from 20.7 to 24.8 per 100 children).

Again, these findings are subject to the same caveats as outlined above for the total sample findings, in that it is not possible to identify whether these seemingly negative Oranga Tamariki-related outcomes are a direct result of the impact of Family Start.

The DiD analyses found a significant increase in the likelihood of Māori children (and/or their siblings) being recorded in a Family Violence notification by Police to Oranga Tamariki in the first year of life by 2.3 per 100 children. The parallel trends test suggested that this outcome could reliably be attributed to the programme. In addition, the DiD analyses found that following the introduction of Family Start, Māori children from the treatment group were 8% (1.9 per 100 children) more likely to be the subject of a Report of Concern in their first year of life. However, in this case the parallel trends test indicated that these effects could not be attributed to the Family Start Programme.[90] All other estimates were not statistically significant.

These findings across both the PSM and DiD analyses were consistent with those obtained by Vaithianathan et al. (2016) for Oranga Tamariki contact and hospitalisations for maltreatment related injuries and long bone fractures. The previous study did not consider Family Violence notifications.

Child protection outcomes: Pasifika

Table 14 presents the PSM- and DiD-estimated impacts of Family Start on Pasifika children. The PSM analyses found significant increases in the likelihood of Oranga Tamariki contact (Reports of Concern,[91] assessments,[92] care placements,[93] and Family Violence notifications[94] by Police to Oranga Tamariki). On the other hand, no significant differences in hospitalisations for maltreatment-related injuries or long bone fractures were recorded.

Consistent with the Māori and total samples, the strongest impacts on Oranga Tamariki contact related outcomes were estimated in the first year of life and decreased in magnitude in subsequent periods. On the other hand, unlike for the Māori and total samples, the PSM analyses did not find any statistically significant differences in the likelihood of being placed into care the first year of life.

The results from the DiD analyses suggested that there were no significant changes following the introduction of the Family Start programme on Pasifika children's likelihood of being the subject of a Report of Concern, assessment, or care placement. There were also no significant differences in the likelihood of hospitalisation for a maltreatment-related injury or a long bone fracture.

---

[90] See Table E5 for more information.

[91] The likelihood of being the subject of a Report of Concern increased in the first year by 96.8% (or 15.0 per 100 children, from 15.5 to 30.5 per 100 children); by 59.8% (or 7.6 per 100 children, from 12.7 to 20.3 per 100 children) in the second year; and by 31.3% (or 3.0 per 100 children, from 9.6 to 12.6 per 100 children) in the sixth year.

[92] The likelihood of being assessed by Oranga Tamariki increased in the first year by 105.4% (or 11.7 per 100 children, from 11.1 to 22.8 per 100 children); by 69.0% (or 5.8 per 100 children, from 8.4 to 14.2 per 100 children) in the second year; and by 30.5% (or 1.8 per 100 children, from 5.9 to 7.7 per 100 children) in the sixth year.

[93] The likelihood of being placed into care increased in the second year by 52.2% (or 1.2 per 100 children, from 2.3 to 3.5 per 100 children); and by 71.4% (or 0.5 per 100 children, from 0.7 to 1.2 per 100 children) in the sixth year. There was no significant difference in first year outcomes.

[94] The likelihood that participants (or their siblings) were recorded in a Family Violence Report of Concern or Contact Record by Police to Oranga Tamariki in the first year increased by 75.2% (11.2 per 100 children, from 14.9 to 26.1 per 100 children); in the second year by 49.7% (7.6 per 100 children, from 15.3 to 22.9 per 100 children); and in the sixth year by 24.0% (3.1 per 100 children, from 12.9 to 16.0 per 100 children).

Table 14. Impact of Family Start on family violence-relevant outcomes (Pasifika)

	Outcome period		
	First year	Second year	Sixth year
Propensity Score Matching estimates			
OT/CYF Report of Concern or other initial intake event	**0.14998***** [0.13531, 0.16465]	**0.07642***** [0.06309, 0.08976]	**0.03031***** [0.01728, 0.04334]
OT/CYF assessment	**0.11722***** [0.10381, 0.13062]	**0.05787***** [0.04638, 0.06936]	**0.01831***** [0.00790, 0.02872]
OT/CYF care placement	0.00326 [-0.00118, 0.00770]	**0.01222***** [0.00567, 0.01877]	**0.00503**** [0.00069, 0.00937]
Child or sibling appeared in a Police FV report to OT/CYF	**0.11184***** [0.09774, 0.12593]	**0.07613***** [0.06251, 0.08976]	**0.03071***** [0.01638, 0.04504]
Hospitalised for a maltreatment related injury	0.00040 [-0.00230, 0.00311]	0.00021 [-0.00297, 0.00338]	-0.00092 [-0.00448, 0.00263]
Hospitalised for a long bone fracture	-0.00039 [-0.00238, 0.00161]		
Difference in Difference estimates			
OT/CYF Report of Concern or other initial intake event	0.00183 [-0.01954, 0.02320]	0.01580 [-0.00866, 0.04026]	-0.00619 [-0.02725, 0.01487]
OT/CYF assessment	0.00259 [-0.01612, 0.02129]	0.00353 [-0.01659, 0.02365]	-0.00716 [-0.02372, 0.00941]
OT/CYF care placement	-0.00006 [-0.00753, 0.00741]	-0.00280 [-0.02081, 0.01520]	-0.00315 [-0.00890, 0.00260]
Child or sibling appeared in a Police FV report to OT/CYF	0.00212 [-0.01826, 0.02251]	-0.00095 [-0.02556, 0.02367]	**-0.02491***** [-0.04430, -0.00551]
Hospitalised for a maltreatment related injury	-0.00341* [-0.00684, 0.00003]	0.00316 [-0.00109, 0.00741]	0.00097 [-0.00720, 0.00914]
Hospitalised for a long bone fracture	-0.00012 [-0.00308, 0.00284]		

Notes: Parameter estimates statistically different from zero at 99% (***), 95% (**), 90% (*) confidence.

The results did indicate a significant reduction in the likelihood of Pasifika children (and/or their siblings) being recorded in a **Family Violence notification** by Police to Oranga Tamariki in the sixth year of life by 2.5 per 100 children. However, a parallel trends test indicated that this impact could not be reliably attributed to Family Start.[95]

These findings across both the PSM and DiD analyses were consistent with those obtained by Vaithianathan et al. (2016) for Oranga Tamariki contact and hospitalisations for maltreatment related injuries and long bone fractures. The previous study did not consider Family Violence notification by Police to Oranga Tamariki.

Child protection outcomes: summary

In summary, PSM analyses indicated a greater likelihood of Family Start participants **interacting with Oranga Tamariki**, and a greater likelihood of participants recording **maltreatment-related hospitalisations**. However, as discussed previously, we cannot determine whether these findings reflect a deterioration of circumstances, reverse causality, and/or safeguarding/surveillance effects. As a result, we cannot determine whether they should be interpreted as positive or negative impacts of the programme, or simply artefacts of the study design.

[95] See Table E5 for more information.

DiD analyses suggested that the introduction of Family Start significantly increases the likelihood of total children and Māori children (and/or their siblings) being recorded in a Family Violence notification by Police to Oranga Tamariki in their first year of life. While the issue of reverse causality does not impact the DiD analyses, the difficulty in assessing whether such findings reflect a real deterioration in circumstances or are instead the result of safeguarding/surveillance effects applies equally to this result. As discussed in previous sections, concerns about the strategy used to identify the target group used for the DiD analyses further limit our confidence in the robustness and reliability of these results.

Given these challenges, we have concluded that we cannot make an evaluative judgement about the impact of Family Start on child protection outcomes in terms of the rubric set out in Appendix A.

DISCUSSION

The results of this impact evaluation provide some promising indications of the positive impact that Family Start is having on the lives and wellbeing of New Zealand's children and their whānau. Most notable were the significant reductions in first year of life mortality, estimated using the PSM approach. These results replicate key findings from Vaithianathan et al. (2016) and suggest that Family Start is having a real impact on children's lives. The fact that our models used a more complex matching process gives us greater confidence in these findings.

The PSM results also found that Family Start participants were more likely to enrol with a PHO, be fully immunised at milestone ages, and attend a B4SC. However, there were mixed results for some outcomes including: no difference in rates of ECE enrolment, an increased likelihood of identifying a significant or non-significant issue at the B4SC and increases in mothers receiving mental health and addiction support or a related benefit. These outcomes are more difficult to interpret, as it is possible that Family Start participation is associated with these outcomes because of case-workers' efforts to ensure that whānau are being connected with services where they might otherwise have gone without support. Further, according to Family Start's Theory of Change, some of the programme's intended short-term outcomes relate to an increase in the *identification* of, and provision of supports for, health and safety issues for children and their whānau. In this respect, we might expect to see an increase in the identification, and therefore records of, health and safety-related issues if Family Start *was* working as intended.

Similarly, PSM outcomes relating to child protection are also difficult to interpret. As discussed above, we cannot determine whether the increased likelihood of engagement with child protection services reflects deteriorating circumstances or indicates increased identification of child-related safety issues by Family Start workers. In addition (especially for year one outcomes), we cannot determine whether the increased likelihood of engagement is a result of programme participation, or the reason why children were referred to the programme (i.e. reverse causality).

It is important to note that the majority of these statistically significant results were identified using the PSM approach. Consistent with Vaithianathan et al. (2016), the results from the DiD analyses largely did not indicate that the programme had an impact on child and whānau outcomes. The only finding that we were able to attribute to the programme using DiD analyses was an increase in the likelihood of children (and/or their siblings) being recorded in a Family Violence notification by Police to Oranga Tamariki (in the first year one life for children of any ethnicity and Māori children).

In our view, these results are likely to reflect the limitations of applying the DiD approach in the context of this evaluation, rather than suggesting that Family Start does not have real impacts on the lives of children and whānau. Family Start has a diverse referral process, which limited our ability to correctly identify which children and whānau would be in the programme's target group. As a result of using a poorly identified target group, the likelihood that DiD analyses would detect statistically significant effects was decreased, particularly when impacts might be small and when outcomes are rare (e.g., mortality). In addition, most of the significant findings from the DiD analyses did not pass the required parallel trends test, further reducing our confidence regarding the suitability of this approach in the present context.

More broadly, the application of the PSM and DiD approaches highlighted some of the limitations of using quasi-experimental methods and administrative data to assess the impact of Family Start on child and whānau wellbeing (particularly for non-Western conceptualisations of wellbeing). This includes challenges in identifying the affected population (and/or the counterfactual), instances where key programme goals cannot be not measured using data held in the IDI, are proxied using administrative interactions, or where the interpretation of findings is unclear (e.g., the impact of potential safeguarding biases). These issues are consistent with some of the challenges discussed in Cram et al. (2018),[96]Wilson et al. (2018) [97] and Matheson (2020).[98] They also highlight the need for the collection of richer, more nuanced data on the wellbeing of individuals and their whānau in New Zealand, to improve researchers' ability to evaluate the outcomes achieved by holistic programmes such as Family Start.

Future analyses of Family Start will need to grapple with whether the data available in the IDI are suitable for assessing a sufficient range of the possible impacts of the programme, particularly for wellbeing of Māori and Pasifika children and whānau. With these caveats in mind, we have identified some analyses that could potentially be explored further in the IDI with the data currently available:

- PSM analyses which characterise Family Start participants relative to their age at programme start rather than at birth, and then matches them with control children whose characteristics are generated the same age. This approach could control for the child's contact with child protection services, and mother's contact with mental health services (among other characteristics) prior to starting Family Start.
- Increase the number of participating children who were linked to the IDI spine. This can be done in instances where even though the FS-Net child record was not able to be linked, their primary caregiver 's record was linked, and the caregiver-child link combination is found in the birth register, and both are linked to the spine. This would materially reduce the number of cases where Family Start participants are incorrectly classified as non-participants in the PSM analysis.
- Exploring the potential to enhance propensity score estimates by incorporating additional information available in the IDI, such as birth parents' income, education, employment and earnings history.
- Completion of PSM analysis of outcomes for children born in the period from 2016 onwards, which might provide some insight into the success of the nationwide expansion of Family Start. Note that analysis of this expansion should ideally be delayed until Family Start is well-established in the new TLAs and until more recent mortality data is available (as Ministry of Health mortality data availability typically lags other data).
- For the DiD approach, potentially the biggest improvement to the current study would be to better identify the Family Start target group. This may be possible if additional data sources were added to IDI in future which better reflected many of the programme referral criteria. Furthermore, future work could explore the use of other approaches such as the synthetic

[96] Cram, F., Vette, M., Wilson, M., Vaithianathan, R., Maloney, T., & Baird, S. (2018). *He awa whiria—braided rivers: Understanding the outcomes from Family Start for Maori*. Evaluation Matters, 165-207. https://www.nzcer.org.nz/system/files/journals/evaluation-maters/downloads/Online_Articles_txt_Cram_FA.pdf

[97] Wilson, M., Hyslop, D., Belgrave, M., Vette, M., and McMillen, M (2018). *Estimating the impact of Social Workers in Schools using linked administrative data.* Ministry of Social Development, Oranga Tamariki.

[98] Matheson, I (2020). *Oranga Tamariki Early Intervention: A synthesis of recent research and evaluations.*

control method to address the many instance where the parallel trends assumption was violated.[99]

Note that these suggestions will not fully address the key limitation of quasi-experimental methods (i.e. unmeasured biases related to programme participation, or 'selection bias'). The key to addressing these limitations is not adding more data, but rather reshaping the measurement system for the intervention (or parts of the intervention) to better demonstrate the impacts of the programme on child and whānau wellbeing. For example, the Early Start programme was piloted using such an approach.[100] However, since the Family Start programme has completed its expansion, potential randomisation would be limited to instances such as the addition of a new service component to the Family Start programme (e.g., by randomising the timing of implementation of the new service at the provider or TLA level).

Overall, the results of the current evaluation tentatively indicate that Family Start is having a positive impact on the wellbeing and safety of children and their whānau in New Zealand. That said, the evaluation also highlights the limitations of using the purely quantitative approaches to evaluate the programme. Given its limitations, it is important to consider the findings from this study alongside the in-depth qualitative process evaluation of Family Start (as in the overall evaluation synthesis report), which adds insights based on the experiences of whānau and providers.

[99] Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association*, 105(*490*), 493-505.

[100] Fergusson, D. M., Horwood, J., Ridder, E., & Grant, H. (2005). *Early Start evaluation report*. Early Start Project Limited. https://www.msd.govt.nz/documents/about-msd-and-our-work/publications-resources/evaluation/early-start-evaluation-report.pdf

Appendix A: ORIGINAL EVALUATION PLAN

Table A1. Desired achievements and performance indicators for the Family Start evaluation

Desired achievement	Performance indicators
Participation in the programme reduces child maltreatment and leads to a safe environment for vulnerable children	• Oranga Tamariki (previously Child Youth and Family) notifications and core assessment phase • Child maltreatment or abuse injuries • All-cause post-neonatal mortality • Post-neonatal injury death • Post-neonatal SIDS/SUDI • FS child or sibling/s had Family Violence Concern Report or CYF intake
Participation in the programme enhances engagement in health and education services	• Enrolment with a Primary Health Organisation (PHO) • Attendance at the Before School Check (B4SC) • Enrolment in early childhood education (ECE) services
Participation in the programme improves child health outcomes	• Childhood immunisations • Children identified through the B4SC as being at a healthy weight • Prevalence of ‘significant issues’ and ‘insignificant issues’ identified at the B4SC
Participation in the programme results in improvements in whānau circumstance[101]	• Mother committing family violence • Mother victim of family violence • Mother involved in criminal justice system • Mother use of mental health or substance use services

[101] When the Evaluation Plan was finalised with input from the EAG, a shared interpretation of the term ‘whānau’ was agreed, limited to the mother, who is assumed to be the primary caregiver of the reference child and thus the whānau member most certain to obtain any benefits of participating in Family Start where the reference child is a Family Start participant.

Table A2. Impact rubric

Criteria	Exceeding expectations	Meeting expectations	Meeting some expectations	Not meeting expectations
Impact	The programme is making a major and positive difference for whānau. Strong evidence exists that the programme is contributing to a safe environment for vulnerable children, and to their mental and physical health. The programme is improving whānau connections skills, circumstances and well-being. No improvements for enhancing impact are required.	The programme is making a positive difference for whānau. The programme is contributing to a safe environment for vulnerable children, and to their mental and physical health. The programme is improving whānau connections, skills, circumstances and well-being. Some small improvements are suggested to enhance programme impact.	The programme is making some difference for whānau. Changes are required to improve safe environments for vulnerable children, enhance their mental and physical health, and further improve whānau connections, skills, circumstances and well-being.	The programme is making little or no difference for vulnerable children and their whānau. Substantial changes are required to improve safe environments for vulnerable children, enhance their mental and physical health, and further improve whānau connections, skills, circumstances and well-being.

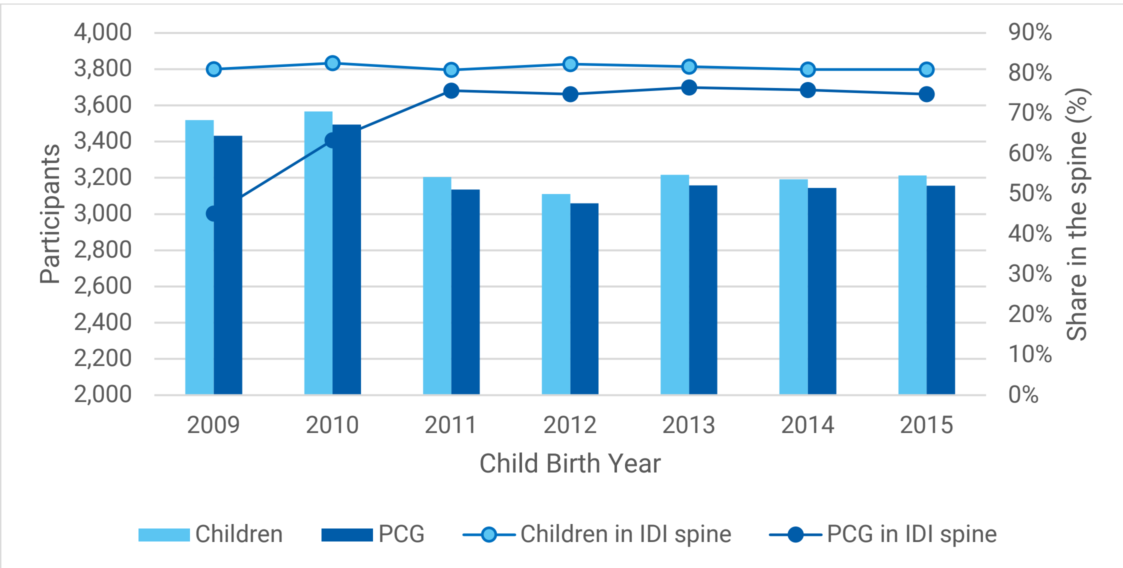
APPENDIX B: FAMILY START PARTICIPANT CHARACTERISTICS

Linking Family Start participants to the IDI spine

FS-NET (the database holding information about all children and their families or whānau enrolled in Family Start) included 36,885 distinct children and 36,183 Primary Care Givers (PCGs). When restricting to 2009-2015 (the period examined under the PSM approach), the number fell to 23,022 children and 22,581 PCGs.

From this cohort, approximately 80% of children and 70% of PCGs were linked to Statistics New Zealand's IDI spine, and were therefore potentially able to be included in the analyses. This was similar to the linkage rate reported by Vaithianathan et al. (2016), where 84% of all children that participated in the programme were linked with other sources (the PCG link rate was not reported). Figure B1 presents the number of children and PCGs that had an active engagement in Family Start, as well as the share of those that were linked to the IDI spine by child's birth year. The figure suggests greater count of participants in 2009-2010, followed by a stable flow of enrolments in the ensuing years.[102] The figure also shows that the linking rate for the IDI was always greater for children than PCGs, and that the linking rate for PCGs was lower in 2009-2010.

Figure B1. Number of children and PCGs actively engaged with Family Start, by year and IDI linkage rate



Source: Organa Tamariki Operational data (2020).

Of the children in the spine, nearly three quarters were linked to a PGC that was also linked to the spine. This share is driven by the overall linkage rate of PCGs, with under half of children linked to a PCG in 2009, nearly two thirds in 2010, and over three quarters in every year that followed. Of the

[102] The greater count in participants born in 2009-2010 possibly reflected the recording of engagement from earlier years as 2009/10.

children linked to a PCG, about 90% showed a link between a child and parent, with this pattern remaining constant over time.

General characteristics of Family Start participant children

The gender split in the programme was fairly even, with 48% of enrolled children being female.[103] Māori accounting for about 60% of children, and Pasifika nearly 30% (note that ethnicity was based on Statistics NZ's standard ethnicity measure [i.e. total ethnicity] rather than prioritised ethnicity, which means that children may be included in both the Māori and Pasifika analyses).

The age of children at the time of joining the programme varied. This may reflect linkage recording and processing complexities. For example, the programme activation age for 5% of children is over 12 months *before* they were born, potentially (partially or fully) reflecting the child being added to an existing engagement (e.g., replacing an older sibling).[104] Approximately 20% of children became active in the programme between 9 months before birth to birth, and approximately 40% of children became active between 1-11 months. Another 20% of children were activated between the age of 1 and 2 years, and 14% at more than two years old.

Family Start referral and completion

In terms of entry pathways to Family Start, over 20% of referrals to the programme were made by the child's parent or other members of their whānau. Other main sources included Well-Child providers/services (17%), other government and non-government agencies (about 17% each) and lead maternity carers (10%).

As shown in Figure B2, on average approximately one third of participating children did not graduate from the programme. This figure is slightly skewed by more recent cohorts, as a greater proportion of children were still active in the programme (whereas those leaving early had already been recorded as 'exited'). When restricting analysis to the 2009-2013 period (where almost all children had either graduated or exited), the graduation rate increases to about 70%. Overall, during this period the average duration of engagement with Family Start for graduates was 5 years 7 months, compared with 3 years 6 months for those exiting the programme prior to graduation.

Predictors of graduation from Family Start

We examined the associations between engagement, child, family, and area factors with graduation for over 11,000 children that were born between 2009 and 2013, participated in Family Start, and either graduated or exited the programme.

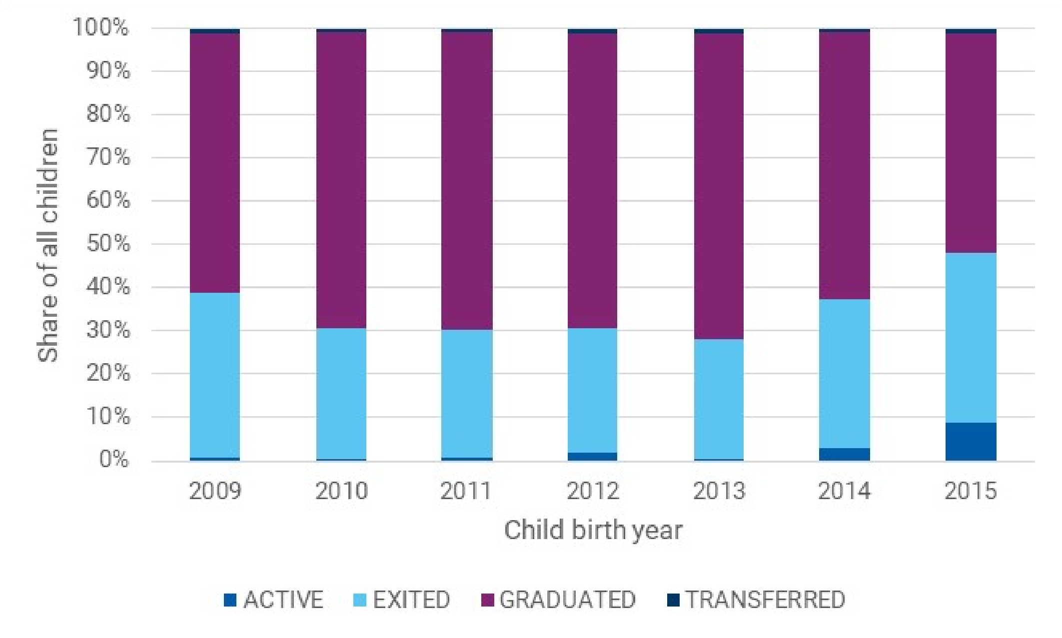
Children that were referred via external providers were almost 14 percentage points more likely to graduate. On the other hand, Māori children (7.6 percentage points), Pasifika children (3.4 percentage points), and children born to a mother receiving a benefit in their first 3 months of life (6.5 percentage points) were less likely to graduate. Transience (residential mobility) was another indicator of lower graduation rates, as was residing in highly deprived meshblocks, and having a mother who was under the age of 25 or single.

[103] Note that non-female observations include a small share of children with unidentified sex.

[104] Only one child can be officially attending the programme at any time.

There was a measurable difference in the graduation rates associated with different referral sources, with the likelihood of graduating increasing depending on the referral source.[105]

Figure B2. Latest status by birth year as at June 2019



Source: Oranga Tamariki operational data (2020).

Engagement by Territorial Local Authority

Until 2015, children born outside TLAs that had Family Start from 2007 or earlier accounted for 2% of the total children actively involved in the programme. This increased to 9% in 2015, 23% in 2016, and 17% in 2017. These data reflect the expansion of the programme into new TLAs in New Zealand in 2016.

Between 2009 and 2015, over one third of children in the programme were born within the Auckland region (largely from Manukau, Waitakere, and Auckland City), followed by 14% in the Waikato Region, and 7-8% in the Northland, Manawatu-Wanganui, and Wellington Regions. While there have been increases in the share for new regions (especially in the South Island), the regions stated above still represent the bulk of participating children. This is expected since these are some of the most populated regions in the country. Canterbury had a lower than representative share, as children in this region are typically referred to the Early Start programme (Early Start was excluded from this evaluation as Early Start data are not available in the IDI).

For 2009-2015, nearly 90% of children resided in the same TLA at birth and during activation in the Family Start programme. When considering TLA grouping, 97% of children that were born in Treated-Until-2015 TLAs remained in this group during Family Start activation. Of participants born in TLAs without Family Start, only half moved to TLAs with the programme during programme activation. This may reflect issues with correctly recording the residential address of children in each time period, and/or the programme providing services outside its official boundaries.

[105] Higher graduation rates were estimated for children referred by General Practitioners, Lead Maternity Carers, Other NGOs, and self/whānau referral compared with other referral sources.

APPENDIX C: OUTCOME AND VARIABLE DEFINITION

Table C1. Outcome definitions

Any OT assessment event
Variable name in code: cyf_ass1/2/6
Detailed description: Reference child is linked to an assessment event at least once in the table cyf_clean.cyf_intakes_event during the relevant age period (0-12 months, 13-24 months, or 60-72 months). Assessment events are either Police Investigations or Reference Child and Family Assessments.

Any OT care placement event
Variable name in code: cyf_care1/2/6
Detailed description: Reference child is linked to a placement event at least once in the table cyf_clean.cyf_placements_event during the relevant age period (0-12 months, 13 months and above, or 60-72 months). Placement events are generated when a placement record is created for an Oranga Tamariki client.

Any OT Report of Concern or other initial intake event
Variable name in code: cyf_intake1/2/6
Detailed description: Reference child is linked to an intake event at least once in the table cyf_clean.cyf_intakes_event during the relevant age period (0-12 months, 13-24 months, or 60-72 months). Intake events include Reports of Concern relating to the Reference child that are received by Oranga Tamariki or the Police, and Youth Justice client intakes which are referred to Oranga Tamariki by the Police, Youth Court or Family Court.

Attended B4SC
Variable name in code: b4sc
Detailed description: Reference child appears in the table moh_clean.b4sc. Age at time of Before School Check attendance was between 36 months and 72 months (combining date of check in the moh_clean.b4sc table with the Reference child's age in the table data.personal_detail). Note B4SC data in the IDI begins in 2011 and is thus unsuitable for DiD analysis which requires data to be collected from 2003.

B4SC identified at least one non-significant issue
Variable name in code: b4sc_unsig_issues
Detailed description: Reference child appears in the table moh_clean.b4sc. Age at time of Before School Check attendance date was between 36 months and 72 months (combining date of check in the moh_clean.b4sc table with the Reference child's age in the table data.personal_detail). Reference child has at least one non-significant issue according to the column moh_bsc_peds_unshaded_nbr. Note B4SC data in the IDI begins in 2011 and is thus unsuitable for DiD analysis which requires data to be collected from 2003.

B4SC identified at least one significant issue
Variable name in code: b4sc_sig_issues
Detailed description: Reference child appears in the table moh_clean.b4sc. Age at time of Before School Check attendance date was between 36 months and 72 months (combining date of check in the moh_clean.b4sc table with the Reference child's age in the table data.personal_detail). Reference child has at least one significant issue according to the column moh_bsc_peds_shaded_nbr. Note B4SC data in the IDI begins in 2011 and is thus unsuitable for DiD analysis which requires data to be collected from 2003.

B4SC record indicated healthy BMI
Variable name in code: b4sc_healthy_bmi
Detailed description: Reference child appears in the table moh_clean.b4sc. Age at time of Before School Check attendance date was between 36 months and 72 months (combining date of check in the moh_clean.b4sc table with the Reference child's age in the table data.personal_detail). B4sc_healthy_bmi is a binary variable = 1 if reference child had a BMI between 18.5 and 25, 0 if Reference child had BMI lower than 18.5 or greater than 25, otherwise = 0. If Reference child participated in a B4SC but weight and/or height were not recorded then BMI cannot be calculated and thus this outcome is coded to 'missing'. Note B4SC data in the IDI begins in 2011 and is thus unsuitable for DiD analysis which requires data to be collected from 2003.

Enrolled with a PHO
Variable name in code: moh_pho1/2/6
Detailed description: Reference child had been enrolled at least once during their lifetime with a Primary Health Organisation according to the table moh_clean.pho_enrolment by the end of the relevant age period (0-12 months, 13-24 months, or 60-72 months).

Enrolled with an ECE provider
Variable name in code: educ_ece
Detailed description: Reference child appears at least once in the table moe_clean.ece_duration (excluding rows which have the classification code 20630 (did not attend) or the duration code 61058 (not regular attendance, only occasionally with no on-going schedule). Note educ_ece does not account for the age of Reference child i.e., whether Reference child was 2 or 10 at the end of the period for which data were available. Measure in the IDI was unsuitable for DiD analysis which requires data to be collected from 2003.

Partially immunised at 1+ milestone age
Variable name in code: moh_immun_part1/2
Detailed description: Reference child was recorded as being up-to-date with immunisations in the table moh_clean.nir atat least one milestone age in the relevant age period (0-12 months or 13-24 months). We did not generate a partial immunisation measure for the 60-72 month period as there was only one milestone age in this age period.

Fully immunised at every milestone age
Variable name in code: moh_immun_full1/2/6
Detailed description: Reference child was recorded as being up-to-date with immunisations

in the table moh_clean.nir at all of the milestone ages in the relevant age period (0-12 months, 13-24 months, or 60-72 months).

Hospitalised for long bone fracture
Variable name in code: moh_lb_inj1
Detailed description: Reference child is linked to at least one hospitalisation event in the table moh_clean.pub_fund_hosp_discharges_event which is linked with one of the following primary diagnosis codes in the table moh_clean.pub_fund_hosp_discharges_diag: Long Bone Injury (S52, S72, S82, T10, T12, S422, S423, S424, S427, S428) at least once during the relevant age period (0-12 months). We were advised by the EAG that it not sensible to analyse long bone injury outcomes over the periods 13-24 months or 60-72 months for physiological reasons so long bone injury data were not analysed for these periods.

Hospitalised for maltreatment related injury
Variable name in code: moh_mal_inj1/2/6
Detailed description: Reference child is linked to a hospitalisation event in the table moh_clean.pub_fund_hosp_discharges_event which is linked with one of the following primary diagnosis codes in the table moh_clean.pub_fund_hosp_discharges_diag: Maltreatment Syndrome (T73, T74, Y06, or Y07), Long Bone Injury (S52, S72, S82, T10, T12, S422, S423, S424, S427, S428), Intracranial Injury (only first year of life) (S06), Assault (X85 – Y09), or **Undetermined cause (Z040, Z045**, Z048) at least once during the relevant age period (0-12 months, 13-24 months, or 60-72 months). Note our code excluded hospitalisations with a primary **diagnosis code of Adverse Circumstances (**Z761, Z865, Z916, Z918) which Oranga Tamariki actuarial code includes – this decision reflected uncertainty about the meaning of the high rate of such hospitalisations in the neonatal period observed in our data which was not observed in Vaithianathan et al.(2016).

Mother received publicly funded mental health services
Variable name in code: m_mhsu
Detailed description: Mother of reference child appeared in any of the following datasets during the relevant period after the birth of the Reference child (0-12 months, 13-24 months, or 25-72 months) : a mental health or substance abuse related record in PRIMHD (moh_clean.PRIMHD), at least one prescription for a substance use or mood related pharmaceutical according to the dataset moh_clean.pharmaceutical, Mental Health or Substance Abuse Public Hospitalisation Discharge Event **(moh_clean.pub_fund_hosp_discharges_diag), and Mental Health or Substance Abuse Lab 'mood' related claims (moh_clean.moh_lab_claims). Measure sourced** from Oranga Tamariki actuarial team.

Post-neonatal injury death
Variable name in code: post_inj1/2/3-6
Detailed description: Reference child appears in the table moh_clean.mortality_registrations with a deceased date that falls within the relevant age period (0-12 months, 13-24 months, or 25-72 months) and which are coded to an ICD-10 primary cause of death code between V01 and Y36 in the column moh_mor_icd_d_code. Deaths at age 0-12 months are only counted if they are also coded 'P' in the column moh_mor_death_type code which indicates post-neonatal mortality (28 days or more after birth). Due to the relative rarity of mortality for older Reference children we chose a broader age period of 25 months - 72 months rather

than simply looking at the period 60-72 months as we have done for non-mortality analyses. Note that, based on the IDI refresh used in this analysis, mortality data were only available until the end of 2015 - data is coded missing to reflect this i.e., Reference children born in 2015 are not included in analyses of post-neonatal injury mortality for any period, Reference children born in 2014 are only included in analyses for the 0-12 month period etc.

Post-neonatal mortality
Variable name in code: post_neo1/2/3-6
Detailed description: Reference child appears in the table moh_clean.mortality_registrations with a deceased date that falls within the relevant age period (0-12 months, 13-24 months, or 25-72 months). Deaths at age 0-12 months are only counted if they are coded 'P' in the column moh_mor_death_type code which indicates post-neonatal mortality (28 days or more after birth). Due to the relative rarity of mortality for older Reference children we chose a broader age period of 25 months - 72 months rather than simply looking at the period 60-72 months as we have done for non-mortality analyses. Note that, for the IDI refresh used in this analysis, mortality data were only available until the end of 2015 – after 2015 data is coded missing to reflect this i.e., Reference children born in 2015 are not included in analyses of post-neonatal mortality for any period, Reference children born in 2014 are only included in analyses for the 0-12 month period etc.

Post-neonatal SUDI
Variable name in code: post_sids1
Detailed description: Reference child appears in the table moh_clean.mortality_registrations with a deceased date that falls within the relevant age period (0-12 months) and which is coded to an ICD-10 primary cause of death code in the following list: R95, R98, R99, W75, W78, W79 in the column moh_mor_icd_d_code. Deaths at age 0-12 months are only counted if they are also coded 'P' in the column moh_mor_death_type code which indicates post-neonatal mortality (28 days or more after birth). Due to the relative rarity of mortality for older reference children we chose a broader age period of 25 months - 72 months as the third period of analysis rather than simply looking at the period 60-72 months as we have done for non-mortality analyses.

Reference child or sibling appeared in Police FV report to OT
Variable name in code: m_fv
Detailed description: Reference child or sibling (linked via mother) appeared in either of the following datasets during the relevant period: Police family violence Report of Concern recorded in the table cyf_clean.cyf_intakes_event or Police family violence contact record recorded in the table cyf_clean.cyf_contact_record. Note that contact records are only recorded to individuals that had an intake event in the Oranga Tamariki system.

Table C2. Variable definitions

	PSM			DiD	
Variable code and description	Exact matching variable	Stage 1 variable	Risk Score component	Target group selection	Controls
birth_dep_index_9_10 Reference child born in NZDep decile 9 or 10 meshblock (2013 NZDEP measure - as per mother's address as recorded in the table data.address_notification)	N	Y	Y	N	Y
birth_nzdep 2013 NZDep score linked to meshblock that reference child's mother lived in a time of birth according to the table data.address_notification. Meshblocks classified using the metadata table clean_read_CLASSIFICATIONS.meshblock_current_higher_geography	N	N	N	Y	Y
child_birth_year Year that reference child was born in as per table data.personal_detail. Note that in DiD analysis cohort birth year quarter was used for controls	Y	Y	N	Y	Y
f_age_at_first_birth Reference child's father's age when first child was born based on data from table data.personal_detail	N	N	N	Y	N
f_ben_3year_pre Mother received a benefit for 3 of the 5 years prior to birth of reference child according to the table msd_clean.msd_spell	N	N	Y	Y	Y
f_ben_4year_pre Mother received a benefit for 4 of the 5 years prior to birth of reference child according to the table msd_clean.msd_spell	N	N	Y	Y	Y

f_birth_year_ind Reference child's father has year of birth information in table data.personal_detail	N	N	N	Y	Y
f_child_count Number of children linked to reference child's father in data.personal_detail table	N	N	N	N	Y
f_corr_comm_pre Reference child's father served a community sentence during the 5 years prior to birth of child according to Corrections data available in the table cor_clean.ov_major_mgmt_periods	N	Y	N	Y	Y
f_corr_cus_pre Reference child's father served a custodial sentence during the 5 years prior to birth of child according to Corrections data available in the table cor_clean.ov_major_mgmt_periods	N	Y	Y	Y	Y
f_corr_pre Reference child's father served a custodial sentence or a community sentence during the 5 years prior to birth of child according to Corrections data available in the table cor_clean.ov_major_mgmt_periods	N	N	Y	Y	Y
f_cyf_care_pre18 Reference child's father had a care placement event before the age of 18 according to CYF/OT data available in the table cyf_clean.cyf_placements_event	N	Y	Y	Y	Y
f_cyf_ink_pre18 Reference child's father had an intake event before the age of 18 according to CYF/OT data available in the table cyf_clean.cyf_intakes_event	N	Y	Y	Y	Y
f_deceased_pre Child's Reference child's father deceased before child's birth (based on deceased date in data.personal_detail table)	N	N	N	Y	Y

f_high_parent_demand Measure derived from Vaithianathan et al. indicates that Reference child's father has 3 or more children including child at time of birth, and/or child was one of multiple birth, and/or child had at least 1 sibling under 2 at time of child's birth. Based on the number of children linked to the child's father in the data.personal_detail table	N	N	N	N	Y
f_incap_nosu_pre5 Reference child's father had at least one spell on an incapacity benefit coded to 'mental disorders' for a reason other than substance abuse during the 5 year period prior to birth of child according to the dataset msd_clean.msd_incapacity	N	Y	Y	Y	Y
f_incap_su_pre5 Reference child's father had at least one spell on an incapacity benefit coded to 'mental disorders' for substance abuse during the 5-year period prior to birth of child according to the dataset msd_clean.msd_incapacity	N	Y	Y	Y	Y
f_known_mshu_pre5 Reference child's father appeared in any of the following datasets during the 5 year period prior to the birth of the child: at least one spell on an incapacity benefit coded to 'mental disorders' according to the dataset msd_clean.msd_incapacity, linked to a mental health or substance abuse related record in PRIMHD (moh_clean.PRIMHD), or had at least one prescription for a substance use or mood related pharmaceutical according to the dataset moh_clean.pharmaceutical. Measure sourced from Oranga Tamariki actuarial team.	N	N	N	N	Y
f_lab_mood_pre5 Reference child's father (according to data.personal_detail table) was linked to at least one lab claim with a test code 'BM2' indicating lithium during the 5 year period prior to the birth of the child, according to the table moh_clean.lab_claims	N	N	N	Y	Y

f_maori Reference child's father (according to the table data.personal_detail) was coded Māori in the table data.personal_detail	N	N	N	N	Y
f_nmds_nosu_pre5 Reference child's father appeared in publicly funded hospital discharge events table moh_clean.pub_fund_hosp_discharges_event with an event that was coded as mental health related (excluding substance abuse codes) in the linked table moh_clean.pub_fund_hosp_discharges_diag	N	N	N	Y	Y
f_nmds_su_pre5 Reference child's father appeared in publicly funded hospital discharge events table moh_clean.pub_fund_hosp_discharges_event with an event that was coded as mental health related and substance abuse related in the linked table moh_clean.pub_fund_hosp_discharges_diag	N	N	N	Y	Y
f_pasifika Reference child's father was coded Pasifika in the table data.personal_detail	N	N	N	N	Y
f_pharma_mood_pre5 Reference child's father had at least one prescription for a mood related pharmaceutical in the five years prior to the birth of the child according to the table moh_clean.pharmaceutical	N	N	N	Y	Y
f_pharma_su_pre5 Reference child's father had at least one prescription for a substance abuse related pharmaceutical in the five years prior to the birth of the child according to the table moh_clean.pharmaceutical	N	N	N	N	Y
f_prihd_mh_pre5 Reference child's father appears at least once in the dataset moh_clean.PRIMHD in the five years prior to the birth of the child linked to an event not coded ‘substance use’ based on OT/MSD actuarial categories	N	N	N	N	Y

f_prihd_su_pre5 Reference child's father appears at least once in the dataset moh_clean.PRIMHD in the five years prior to the birth of the child linked to an event coded 'substance use' based on OT/MSD actuarial categories	N	N	N	N	Y
f_sib_cyf_care_pre Sibling of reference child (linked via father using the table data.personal_detail) had a care placement event during the 5 year period prior to the birth of the child according to the dataset cyf_clean.cyf_placements_event	N	Y	N	Y	Y
f_sib_cyf_ink_pre Sibling of reference child (linked via father using the table data.personal_detail) had an intake event during the 5 year period prior to the birth of the child according to the dataset cyf_clean.cyf_intakes_event	N	Y	N	Y	Y
f_sib_fvcr_pre Sibling of reference child (linked via father using the table data.personal_detail) appears in the table cyf_clean.cyf_contact_record_events with a contact record with a family violence indicator during the 5 year period prior to the birth of the child	N	N	Y	Y	Y
f_sib_fvcr_pre1 Sibling of reference child (linked via father using the table data.personal_detail) appears in the table cyf_clean.cyf_contact_record_events with a contact record with a family violence indicator during the 1 year period prior to the birth of the child	N	N	Y	Y	Y
f_sib_fvink_pre Sibling of reference child (linked via father using the table data.personal_detail) appears in the table cyf_clean.intakes_details with a contact record with a family violence indicator during the 5 year period prior to the birth of the reference child	N	N	Y	Y	Y
f_sib_fvink_pre1 Sibling of reference child (linked via father using the table data.personal_detail)	N	N	Y	Y	Y

appears in the table cyf_clean.intakes_details with a contact record with a family violence indicator during the 1 year period prior to the birth of the reference child					
f_teen Father of reference child (as per table data.personal_detail) was under 18 when reference child was born based on father's age in data.personal_detail table	N	N	N	N	Y
f_under_20 Father of reference child (as per table data.personal_detail) was under 20 when reference child was born based on father's age in data.personal_detail table	N	N	N	N	Y
f_under_25 Father of reference child (as per table data.personal_detail) was under 25 when reference child was born based on father's age in data.personal_detail table	N	N	N	N	Y
father Reference child linked to a father record in table data.personal_detail	N	Y	N	Y	Y
female Reference child is female according to the table data.personal_detail	Y	Y	N	Y	Y
low_birth_weight Reference child had low birth weight (lower than 2500 grams) according to the table moh_matb_birthweight_nbr	N	Y	N	N	N
m_age_at_first_birth Reference child's mother's age when first reference child was born based on data from table data.personal_detail	N	N	N	Y	N
m_ben_3year_pre Mother received a benefit for 3 of the 5 years prior to birth of reference child according to the table msd_clean.msd_spell or had a partner who did	N	Y	N	Y	Y

m_ben_4year_pre Mother received a benefit for 4 of the 5 years prior to birth of reference child according to the table msd_clean.msd_spell or had a partner who did	N	N	N	Y	Y
m_ben_age_combo Three part variable derived from Vaithianathan et al. 2016 which categorises reference children as follows: 1. reference child's mother (as per table data.personal_detail) was aged under 20 at time of reference child's birth 2. reference child's mother was over 20 at time of reference child's birth and received a benefit during the thirteen weeks after the reference child was born (or partner did). 3. reference child's mother was over 20 at the time of the reference child's birth and did not receive a benefit during the 13 weeks after reference child's birth (and was not linked to a partner who did).	Y	N	N	Y	N
m_child_count Number of children linked to reference child's mother in the data.personal_detail table	N	N	N	Y	Y
m_corr_comm_pre Reference child's mother served a community sentence during the 5 years prior to birth of child according to Corrections data available in the table cor_clean.ov_major_mgmt_periods	N	Y	N	Y	Y
m_corr_cus_pre Reference child's mother served a custodial sentence during the 5 years prior to birth of child according to Corrections data available in the table cor_clean.ov_major_mgmt_periods	N	Y	N	Y	Y
m_corr_pre Reference child's mother served a custodial sentence or a community sentence during the 5 years prior to birth of child according to Corrections data available in the table cor_clean.ov_major_mgmt_periods	N	N	N	Y	Y

m_cyf_care_pre18 Reference child's mother had a care placement event before the age of 18 according to CYF/OT data available in the table cyf_clean.cyf_placements_event	N	Y	N	Y	Y
m_cyf_ink_pre18 Reference child's mother had an intake event before the age of 18 according to CYF/OT data available in the table cyf_clean.cyf_intakes_event	N	Y	N	Y	Y
m_fs_child_first Reference Child was mother's first child (earliest birth date among children linked to child's mother via the table data.personal_detail)	N	Y	N	N	N
m_high_parent_demand Measure derived from Vaithianathan et al. indicates that, at time of reference child's birth, mother is linked to 3 or more total children under 18 (including child) , or mother is linked to another child aged under 2. Child's mother linked to child's siblings via table data.personal_detail	N	N	Y	N	N
m_incap_nosu_pre5 Reference child's mother had at least one spell on an incapacity benefit coded to 'mental disorders' for a reason other than substance abuse during the 5-year period prior to birth of child according to the dataset msd_clean.msd_incapacity	N	N	Y	Y	Y
m_incap_su_pre5 Reference child's mother had at least one spell on an incapacity benefit coded to 'mental disorders' for substance abuse during the 5-year period prior to birth of child according to the dataset msd_clean.msd_incapacity	N	N	Y	N	Y
m_known_mhsu_5_pre Reference child's mother appeared in any of the following datasets during the 5 year period prior to the birth of the child: at least one spell on an incapacity benefit coded to 'mental disorders' during the 5 year period prior to birth of child according to the dataset msd_clean.msd_incapacity, a mental health or substance abuse related record in PRIMHD (moh_clean.PRIMHD), or had at least one prescription for a substance use or mood related pharmaceutical according	N	Y	Y	N	Y

to the dataset moh_clean.pharmaceutical. Measure sourced from Oranga Tamariki actuarial team.					
m_lab_mood_pre5 Reference child's mother (according to data.personal_detail table) was linked to at least one lab claim with a test code 'BM2' indicating lithium during the 5 year period prior to the birth of the child, according to the table moh_clean.lab_claims	N	N	N	Y	Y
m_maori Reference child's mother (according to data.personal_detail table) was coded Māori in the table data.personal_detail	N	N	N	Y	Y
m_nmds_nosu_pre5 Reference child's mother (according to data.personal_detail table) appeared in publicly funded hospital discharge events table moh_clean.pub_fund_hosp_discharges_event with an event that was coded as mental health related (excluding substance abuse codes) in the linked table moh_clean.pub_fund_hosp_discharges_diag	N	N	N	Y	Y
m_nmds_su_pre5 Reference child's mother (according to data.personal_detail table) appeared in publicly funded hospital discharge events table moh_clean.pub_fund_hosp_discharges_event with an event that was coded as mental health related and substance abuse related in the linked table moh_clean.pub_fund_hosp_discharges_diag	N	N	N	Y	Y
m_pasifika Reference child's mother (according to data.personal_detail table) was coded Pasifika in the table data.personal_detail	N	N	N	Y	Y
m_pharma_mood_pre5 Reference child's mother had at least one prescription for a mood related pharmaceutical in the five years prior to the birth of the reference child according to the table moh_clean.pharmaceutical	N	N	Y	N	Y

m_pharma_su_pre_5 Reference child's mother had at least one prescription for a substance abuse related pharmaceutical in the five years prior to the birth of the reference child according to the table moh_clean.pharmaceutical	N	N	Y	N	Y
m_primhd_mh_pre5 Reference child's mother appears at least once in the dataset moh_clean.PRIMHD in the five years prior to the birth of the reference child linked to an event not coded 'substance use' based on OT/MSD actuarial categories	N	N	Y	N	Y
m_primhd_su_pre5 Reference child's mother appears at least once in the dataset moh_clean.PRIMHD in the five years prior to the birth of the reference child linked to an event coded 'substance use' based on OT/MSD actuarial categories	N	N	Y	N	Y
m_sib_cyf_care_pre Sibling of reference child (linked via mother using the table data.personal_detail) had a care placement event during the 5 year period prior to the birth of the reference child according to the dataset cyf_clean.cyf_placements_event	N	Y	Y	Y	Y
m_sib_cyf_ink_pre Sibling of reference child (linked via mother using the table data.personal_detail) had an intake event during the 5 year period prior to the birth of the reference child according to the dataset cyf_clean.cyf_intakes_event	N	Y	N	Y	Y
m_sib_fvcr_pre Sibling of reference child (linked via mother using the table data.personal_detail) appears in the table cyf_clean.cyf_contact_record_events with a contact record with a family violence indicator during the 5 year period prior to the birth of the reference child	N	N	N	Y	Y
m_sib_fvcr_pre1 Sibling of reference child (linked via mother using the table data.personal_detail) appears in the table cyf_clean.cyf_contact_record_events with a contact record	N	N	N	Y	Y

with a family violence indicator during the 1 year period prior to the birth of the reference child					
m_sib_fvink_pre Sibling of reference child (linked via mother using the table data.personal_detail) appears in the table cyf_clean.intakes_details with a contact record with a family violence indicator during the 5 year period prior to the birth of the reference child	N	N	N	Y	Y
m_sib_fvink_pre1 Sibling of reference child (linked via mother using the table data.personal_detail) appears in the table cyf_clean.intakes_details with a contact record with a family violence indicator during the 1 year period prior to the birth of the reference child	N	N	N	Y	Y
m_smoke_birth Mother of reference child (as per table data.personal_detail) smoked during pregnancy prior to birth of reference child according to information provided by primary care provider recorded in MoH maternity data in the table moh_clean.maternity_mother.	N	Y	N	N	Y
maori Reference child was coded Māori in the table data.personal_detail. Note that this is an 'ever Māori' variable which reflects multiple administrative data sources within the IDI.	Y	Y	N	Y	Y
mb_count Number of distinct meshblocks reference child's mother lived in during the period from 6 month prior to birth of reference child to 1.5 years post-birth based on mother's addresses in data.adresss_notification table at time of reference child's birth	N	N	N	Y	Y
mother_on_ben Mother of reference child (as per table data.personal_detail) or mother's partner received a benefit during the first 13 weeks after reference child's birth -	N	Y	N	Y	Y

according to the table msd_clean.msd_spell. This included mother being either the main or secondary applicant					
mother_single Reference child's mother defined as single at birth of reference child if any of the following criteria are met: 1. parental relationship is coded to 'no' in the column dia_parents_rel_code in the table dia_clean.births or 2. reference child's mother's partner does not have snz_uid record in the table dia_clean.births or 3. reference child's mother received sole parent or widow's benefit in the first three months after the reference child's birth	N	Y	Y	Y	Y
mother_teen_parent_fs Mother of reference child (as per table data.personal_detail) was under 18 when reference child was born based on mother's age in data.personal_detail table	N	N	Y	N	N
mother_under_20 Mother of reference child (as per table data.personal_detail) was under 20 when reference child was born based on mother's age in data.personal_detail table	N	Y	Y	N	Y
mother_under_25 Mother of reference child (as per table data.personal_detail) was under 25 when reference child was born based on mother's age in data.personal_detail table	N	Y	Y	Y	Y
nzdep_25 Reference child was born in a meshblock in the top deprivation quartile (based on 2013 NZDEP scores) based on mother's address in the table data.address_notifcation at the time of reference child's birth. Meshblock level NZDEP scores are based on the 2013 results, obtained from metadata table meshblock_current_higher_geography	N	N	N	Y	Y
nzdep_mscore The mean NZDEP score (2013 scores) of the meshblocks reference child's mother resided in from 6 months prior to reference child's birth to 1.5 years post-	N	N	N	Y	Y

birth according to the table data.address_notification. Meshblock level NZDEP scores obtained from metadata table meshblock_current_higher_geography					
pasifika Reference child was coded Pasifika in the table data.personal_detail	N	Y	N	Y	Y
pasifika_prioritised Reference child was coded Pasifika and not Māori in the table data.personal_detail (prioritisation used for matching purposes - separate outcomes for Pasifika reference children and families do not use prioritised ethnicity)	Y	N	N	N	N
parent_count Number of parents linked to reference child in table data.personal_detail	N	N	N	N	Y
post1 Reference child was born in a TLA where Family Start was available based on mother's address meshblock in data.adresss_notification table at time of reference child's birth. Meshblocks classified using 2010 Territorial Authority Classifications. Data on Family Start availability sourced from previous evaluation and also direct consultation with Oranga Tamariki staff	N	N	N	N	Y
risk_score The use of a risk score as a component of PSM and DiD matching methods was informed by the previous evaluation. Risk score sums binary characteristics of reference child across the following variables: birth_dep_index_9_10, f_ben_3year_pre, f_ben_4year_pre, f_corr_pre, f_cyf_care_pre18, f_cyf_ink_pre5, f_incp_su_pre5, f_incp_nosu_pre5, f_sib_fvcr_pre, f_sib_fvcr_pre1, f_sib_fvink_pre, f_sib_fvink_pre1, m_corr_pre, m_high_parent_demand, m_incap_nosu_pre5, m_incap_su_pre5, m_known_mhsu_pre5, m_pharma_mood_pre5, m_pharma_su_pre5, m_primhd_mh_pre5, m_primhd_su_pre5, m_sib_cyf_care_pre, mother_teen_parent_fs, mother_under_20, mother_under_25, mother_single	N	Y	N	Y	Y

ta_code The numeric code identifying the TLA that reference child was born in based on mother's address meshblock in data.adresss_notification table at time of reference child's birth. Meshblocks classified using 2010 Territorial Authority Classifications and linked using the table clean_read_CLASSIFICATIONS.meshblock_current_higher_geography	N	N	N	Y	Y
ta_count Number of Territorial Local Authorities reference child's mother lived in during the period from 6 month prior to birth of reference child to 1.5 years post-birth based on mother's addresses in data.adresss_notification table. Meshblocks classified using 2010 Territorial Authority Classifications and linked using the table clean_read_CLASSIFICATIONS.meshblock_current_higher_geography	N	N	N	N	Y
ta_more1 Reference child's mother resided in more than one Territorial Local Authority during the period from 6 months before reference child's birth to 1.5 years post-birth based on address meshblock data recorded in the table data.address_notification. Meshblocks classified using 2010 Territorial Authority Classifications and linked using the table clean_read_CLASSIFICATIONS.meshblock_current_higher_geography	N	N	N	N	Y
teen_share Proportion of children born to mothers under the age of 18. Data aggregated at the Territorial Authority and birth-year/quarter level. Data for mother's age of is sourced from data.personal_detail table	N	N	N	N	Y
urban The proportion of days the reference child's mother spent living in meshblocks coded 'Urban Areas' as per the IDI metadata table clean_read_CLASSIFICATIONS.meshblock_current_higher_geography during the period from 6 months prior to child's birth to 1.5 years post-birth (as per the mother's addresses as recorded in table data.address_notification)	N	N	N	Y	Y

urban_birth_2 Reference Child born in meshblock coded 'major urban area' (as per mother's address as recorded in the table data.address_notification linked with IDI meshblock metadata stored in the table clean_read_CLASSIFICATIONS.meshblock_current_higher_geography) OR reference child coded Pasifika in the table data.personal_detail. The Pasifika child exception reflects the high concentration of Pasifika people in major urban areas which made matching of Pasifika children not born in major urban areas problematic.	Y	Y	N	Y	Y

APPENDIX D: ADDITIONAL PSM ANALYSES

Replication of the PSM analysis in the original 2016 study

One of the objectives of this evaluation was to explore the extent to which Vaithianathan et al.'s (2016) results could be replicated. For this reason, we implemented a specification that was as close as possible to their original specification. Having done this, we made a number of changes to address some of the issues and limitations we identified.

The method applied in Vaithianathan et al. (2016) involved two stages. Stage one involved estimating a propensity score model for children who lived in areas where Family Start was available at the time of their birth. The propensity score model was generated using a standard logit regression model, as shown below. The dependent variable Y indicates Family Start participation, the independent variables $(Z_1, \ldots, Z_j)$ included in the model are essentially the same predictors of programme participation used by Vaithianathan et al. (see Appendix C of this report for a complete list of variables) and f represents the cumulative logistic distribution function.

$$\Pr\left(Y = 1 \middle| Z_1, \ldots, Z_j\right) = f(\beta_0 + \beta_1 Z_1 + \cdots + \beta_j Z_j)$$

The predictors of participation were determined at the time the reference child was born, for example whether or not any siblings had contact with child-protection services before the child's birth. The only exception to this was whether the child's mother was supported by a main benefit within 13 weeks following birth.

Following the initial stage one regression, the probability of participation was estimated for all children in the dataset (including for children born in TLAs where Family Start was not available). Then, following Vaithianathan et al. (2016), control individuals were selected from TLAs where Family Start was not available.

Stage two involved matching participant children with suitable non-participant children based on stage one propensity scores and some exact matching criteria (these were as close as possible to those used originally):

- whether the child was supported by a main benefit within 13 weeks following birth,[106] combined with mother's age (3 categories: 1. mother under 20, 2. mother 20 and over and receiving main benefit during child's first 13 weeks of life, 3. mother 20 and over and did not receive main benefit during child's first 13 weeks of life)
- living in a neighbourhood at birth that was in NZDep 9 or 10 (the most deprived quintile)
- ethnicity of the child (Māori, Pasifika if not Māori)
- urban location combined with Pasifika ethnicity

[106] Limitation in data availability at the time of the study resulted in using mother benefit status (rather than child's), and covering a 3 month period rather than 13 weeks.

- birth year.

Following Vaithianathan et al. (2016), we implemented the matching using the STATA command *teffects nnmatch*. The method combined 1-1 nearest neighbour matching on the stage one propensity score with exact matching on selected characteristics (with all ties retained). The Average Effect of Treatment (ATE) was estimated, having excluded individuals with a propensity score lower than the stage one median score from the analysis (nearly all individuals dropped were potential controls).

Vaithianathan et al. (2016) focussed on participant children born from 2009 to 2011, in TLAs where Family Start became available during 2005 to 2007 (referred to as the '2005-2007 expansion TLAs'). In our replication of their analysis there were 3,297 participants and 29,865 individuals in the never-treated TLAs (potential controls). Estimating Average Treatment Effect (ATE) rather than the Average Treatment effect on the Treated (ATT) means that in addition to all participants being matched to at least one control, all controls are matched to a least one participant. This leads to a much larger total sample size of 33,162 children and to more precisely estimated impacts.

Table D1 compares the characteristics of the matched sample in our replication analysis. There were some significant differences between matched treatments and controls; in particular, the mothers of matched participants were more likely than matched controls to be single at the child's birth, to be first time mothers, and (to a lesser extent), to have been supported by benefits for at least three of the last five years. This suggests that the mothers of matched participants may have been a slightly higher at-risk group than the mothers of matched controls (thereby further confirming the decision to remove comparison of maternal outcomes from the outcomes assessed in the current analyses).

Table D2 compares the estimated impacts with those obtained in the original study. Estimated impacts were very similar, with statistically significant reductions in post-neonatal mortality in the first year, and deaths from SUDI and injury. Estimated impacts on immunisation rates and contact with Oranga Tamariki (CYF) were also very similar. The original study found a significant reduction in the likelihood participants were enrolled with a PHO, whereas we find no significant impact.

Table D1. Comparison of matched participants and controls, 2021 replication (2009-2011 births)

	Matched participants	Matched comparisons	Difference
Number of participants (weighted)	33,162	33,162	
Māori	0.4105	0.4105	0.0000
Pasifika	0.1409	0.1425	-0.0016
Pasifika and not Māori	0.0833	0.0833	0.0000
Female	0.4226	0.4698	-0.0472
Child had a low birth weight	0.0804	0.0846	-0.0042
Child was born in NZDEP 9-10 area	0.2887	0.2887	0.0000
Child was born in a major urban area	0.3399	0.2604	0.0795
Child was born in a major urban area or was of Pacific ethnicity	0.3339	0.3339	0.0000
Child is their mother's first born	0.4984	0.4404	0.0579
Mother			
Single at the birth of the child	0.5103	0.4269	0.0833
Supported by benefit within 3 months of child's birth	0.4453	0.4426	0.0027
Under 18 when child was born	0.0236	0.0235	0.0002
Under 20 when child was born	0.1270	0.1270	0.0000
Under 25 when child was born	0.4704	0.4641	0.0063
Smoked at the time of the child's birth	0.2194	0.2445	-0.0251
Supported by benefits for 3 or more of the last 5 years	0.1767	0.1538	0.0229
Received mental health services or prescription in the 5 years before the child's birth	0.1662	0.1823	-0.0161
Had a child protection placement before age 18	0.0460	0.0446	0.0014
Had a child protection notification before age 18	0.2269	0.2342	-0.0073
Sibling[107] had a child protection placement during the 5 years prior to child's birth	0.0209	0.0155	0.0054
Sibling had a child protection notification during the 5 years prior to child's birth	0.1872	0.1652	0.0220
Served a community service sentence during the 5 years prior to child's birth	0.0662	0.0719	-0.0057
Served a custodial sentence during the 5 years prior to child's birth	0.0091	0.0111	-0.0020
Father			
Recorded on birth certificate and linked to IDI spine	0.9025	0.9227	-0.0202
Received an incapacity benefit due to substance abuse in 5 years prior to child's birth	0.0189	0.0209	-0.0020
Received an incapacity benefit due to mental health in 5 years prior to child's birth	0.0437	0.0589	-0.0152
Had a child protection placement before age 18	0.0335	0.0351	-0.0016
Had a child protection notification before age 18	0.1622	0.1858	-0.0236
Served a community service sentence during the 5 years prior to child's birth	0.2001	0.1934	0.0067
Served a custodial sentence during the 5 years prior to child's birth	0.0789	0.0757	0.0031
Risk count (child)	3.1434	3.1480	-0.0046
Propensity score (child)	0.0926	0.0919	0.0007

Notes: Underlying counts are rounded in accordance with Stats NZ's random rounding 3 regulations. Percentages derived from fewer than 6 observations have been supressed ('s'). For full list of variables, see Appendix C.

[107] **Siblings refers to all other children linked to the child's mother.**

Table D2. PSM - Vaithianathan et al. (2016) and replication (2021)

	Vaithianathan et al. (2016) Estimate [95% CI]			Replication (2021) Estimate [95% CI]		
	Outcomes in the first year	Outcomes in the second year	Outcomes in the fifth year	Outcomes in the first year	Outcomes in the second year	Outcomes in the fifth year
Number of observations	33,087	33,087	11,029	33,162	33,162	33,162
Post neonatal mortality	**-0.0016*** **	-0.0003*		**-0.0015** **	-0.0003	
	[-0.0026, -0.0005]	[-0.001, 0.000]		[-0.0029, -0.0000]	[-0.0007, 0.0001]	
Post neonatal SUDI	**-0.0007*** **	-0.0001*		**-0.0006** **	**-0.0002** **	
	[-0.0011, -0.0004]	[0.000, 0.000]		[-0.0012, -0.0000]	[-0.0003, -0.0000]	
Post neonatal SUDI	**-0.0011*** **			**-0.0010*** **		
	[-0.0016, -0.0006]			[-0.0017, -0.0004]		
OT/CYF Report of Concern or other initial intake event	**0.1751*** **	**0.1090*** **		**0.1609*** **	**0.0748** **	
	[0.125, 0.225]	[0.077, 0.141]		[0.0811, 0.2408]	[0.0025, 0.1472]	
OT/CYF care placement	0.0063*	0.0051*		0.0026	**0.0252*** **	
	[-0.001, 0.014]	[0.000, 0.010]		[-0.0128, 0.0180]	[0.0074, 0.0430]	
Hospitalised for a maltreatment related injury	0.0002	0.0011		0.0034*	0.0026	
	[-0.003, 0.003]	[-0.001, 0.003]		[-0.0006, 0.0073]	[-0.0032, 0.0083]	
Hospitalised for a long bone fracture	0.0025			0.0000		
	[-0.001, 0.006]			[-0.0024, 0.0024]		
Enrolled with a PHO	**-0.0306*** **	0.0020		-0.0005	0.0025	
	[-0.048, -0.013]	[-0.009, 0.013]		[-0.0083, 0.0073]	[-0.0015, 0.0065]	
Fully immunised at 1+ milestone age	**0.0499*** **	**0.0375*** **		0.0252	0.0273	
	[0.027, 0.072]	[0.018, 0.057]		[-0.0088, 0.0592]	[-0.0053, 0.0599]	
Fully immunised at every milestone age	0.0486*	0.0543*		0.0478*	0.0371*	
	[-0.009, 0.106]	[-0.002, 0.111]		[-0.0002, 0.0957]	[-0.0041, 0.0782]	
Enrolled with an ECE provider		*2009 cohort*	**0.0755*** **		*2009-11 cohorts*	-0.0504
			[0.033, 0.118]			[-0.1304, 0.0295]
Attended B4SC		*2009 cohort*	-0.1364		*2009-11 cohorts*	-0.0190
			[-0.395, 0.123]			[-0.0877, 0.0497]

Notes: Parameter estimates statistically different from zero at 99% (***), 95% (**), 90% (*) confidence. Counts are rounded with accordance to Stats NZ's random rounding 3 regulations. For full list of variables, see Appendix C

Impact of changes to the PSM method on the results obtained for the 2009-2011 replication cohort

Adjustments to the modelling approach

Having implemented a PSM specification that was as close as possible to Vaithianathan et al.'s specification, and having demonstrated that we were able to produce comparable estimates, we made some changes to how PSM was implemented in the current evaluation in order to address some of the limitations we identified with the previous analysis:

- estimation of the Average Treatment Effect (ATE) rather than standard approach of estimating the Average Treatment Effect on Treated (ATT)
- *teffects* ignores the calliper option when exact matching is included and requires all individuals be matched to the specified number of neighbours, leading to matches being made between individuals with very different propensity scores in some cases
- selecting controls from TLAs where Family Start was not available resulted in some significant differences between the average characteristics of matched treatments and controls, as noted above. In addition, it increases the risk of area-level differences between the treatment and control TLAs affecting the observed differences in outcomes
- matching was done based on characteristics at the time of the child's birth rather than at the time they were referred or started Family Start. We considered this to be particularly problematic for the estimation of impacts on mortality rates.[108]

We made the changes described in PSM methods section in the main body of this report. Participants were matched to non-participants from TLAs where Family Start was available. *psmatch2* was used to implement a calliper of 0.015 and ATT was estimated. We also modified the matching approach to ensure that each participant was only matched to children who were alive at the age (in months) that the participant started Family Start.

Impact of the changes on results obtained

Table D3 shows the impact of implementing our primary specification on the results obtained for the 'replication cohort', that is, participant children born in 2009 to 2011, in TLAs where Family Start became available during 2005 – 2007.

Column 1 corresponds to estimates obtained from our replication of the method used by Vaithianathan et al. (2016); Column 2 to the estimation of ATT rather than ATE using the *teffects* specification; Column 3 to using *psmatch2* to estimate ATT and selecting controls from TLAs where Family Start was not available; Column 4 to using the same approach as for Column 3, but with the modified matching method for mortality outcomes whereby participants are matched to

[108] About 22% of child participants were enrolled in Family Start before their birth, 34% between birth and before six months, 17% between six months and before 12 months, and 17% after 12 months. It is likely that in some cases contact with Heath and child protection services in the early months of a child's life may have resulted in or influenced a referral to Family Start; this creates a confound given that these are also measured outcomes. For non-mortality outcomes, Vaithianathan et al. (2016) re-estimated the second-year outcomes for the sub-sample that did not record that same outcome in the first year (e.g., mental health service use in year 2 was estimated for those who did not have a recorded mental health service use in year 1). Overall, they did not find significant differences for this sub-sample.

controls who were still alive when the child entered Family Start; Column 5 selects controls from TLAs where Family Start became available in 2005 – 2007 (that is, the same TLAs as participants); and Column 6 to the same approach as for Column 5, but with the modified matching method for mortality outcomes.

The main impact of estimating ATT rather than ATE was the much reduced number of matched controls, wider confidence intervals, and reduced number of significant effects. The increases in immunisation rates and contacts with child protection services were still significant, but the reductions in mortality in the first year were no longer significant.

Comparing results from the different ATT specifications, we see that the impacts vary depending on the outcome being considered. Note that the mean outcomes for matched participants do not change across columns, reflecting the fact that the participant sample was essentially the same across the different specifications, but the composition of the control group did change (and potentially with it the mean outcome for controls and the estimated impact).

Selecting controls from the same TLAs as participant children, rather than from TLAs where Family Start was not available, meant that differences in outcomes by TLA are controlled for. For example, comparing Columns 3 and 5, we see that a significant reduction in ECE participation rates of 2.8 percentage points becomes a non-significant increase of 1.4 percentage points, and the small non-significant reduction in PHO enrolment rates in the first year becomes a significant 1.7 percentage point increase. The estimated impacts on immunisation rates are unchanged and significantly positive. The estimated impact on SUDI deaths is reduced from 1.9 per 1000 children (95%CI, 0.05 to 3.7) significant at the 5% level, to 0.44 per 1000 children (95%CI, -0.43 to 1.3), while the impact on post-neonatal mortality in the first year increases from 1.6 per 1000 children (95%CI, -0.41 to 3.6) not significant at the 10% level, to 2.3 per 1000 children (95%CI, 0.61 to 4.1) significant at the 1% level.

Implementing the modified matching method for mortality outcomes reduces the estimated impact on post-neonatal infant mortality in the first year from 2.3 per 1000 children (95%CI, 0.61 to 4.1) significant at the 1% level, to 1.1 per 1000 children (95%CI, -0.17 to 2.4) significant at the 10% level. The reduction in injury deaths in the first year of 0.69 per 1000 children (95%CI, 0.03 to 1.34) significant at the 5% level, is reduced to 0.55 per 1000 children (95%CI, 0.15 to 0.95) significant at the 1% level. The reduction in SUDI deaths in the first year of 0.44 per 1000 children (95%CI, -0.43 to 1.3) is reduced to 0.12 per 1000 children (95%CI, -0.51 to 0.76).

Comparing result from our preferred models for births during the 2009-2011 period (Column 6 for mortality outcomes and Column 5 for all other outcomes) with those obtained by Vaithianathan et al. (2016), we find similar significant increases in immunisation rates and contact with child protection services. We find a significant increase in PHO enrolment in the first year, which differs from the earlier study which found a significant decline. Our estimated mortality impacts are smaller than those obtained in the earlier study. We find only the reductions in injury deaths in the first and second year to be statistically significant at the 5% level and the reduction in post-neonatal mortality in the first year to be significant at the 10% level. The original study found a significant reduction at the 1% confidence level in post-neonatal mortality, SUDI and injury deaths in the first year, and significant reduction at the 10% level in both post-neonatal mortality and injury deaths in the second year.

Table D3. Comparison of PSM results for various specifications (2009-2011 births)

	teffect nnmatch		psmatch2			
Comparison *Estimation* *Adjustment*	Outside FS TLAs ATE (1)	Outside FS TLAs ATT (2)	Outside FS TLAs ATT (3)	Outside FS TLAs ATT matching adj (4)	Inside FS TLAs ATT (5)	Inside FS TLAs ATT matching adj (6)
Number of Participants	3,297	3,297	3,258	3,258	3,282	3,282
Number of Comparisons	29,865	4,986	6,669	6,639	9,612	9,591
Weighted participants	33,162	3,297	3,258	3,258	3,282	3,282
Weighted comparisons	33,162	3,297	3,258	3,258	3,282	3,282
Post neonatal mortality in first year	**-0.00146****	-0.00131	-0.00161	-0.00068	**-0.00233*****	-0.00110*
	[-0.0029, 0.0000]	[-0.0062, 0.0036]	[-0.00363, 0.00041]	[-0.00199, 0.00063]	[-0.00406, -0.00061]	[-0.00236, 0.00017]
Post neonatal mortality in second year	-0.00030	-0.00031	-0.00031	-0.00025	-0.00018	-0.00006
	[-0.00068, 0.00008]	[-0.0017, 0.0011]	[-0.00114, 0.00052]	[-0.00090, 0.00041]	[-0.00152, 0.00115]	[-0.00096, 0.00084]
Post neonatal injury death in first year	**-0.00058****	-0.00101	**-0.00143***	**-0.00109*****	**-0.00069****	**-0.00055*****
	[-0.00115, -0.00001]	[-0.0056, 0.0036]	[-0.00291, 0.00005]	[-0.00185, -0.00033]	[-0.00134, -0.00003]	[-0.00095, -0.00015]
Post neonatal injury death in second year	**-0.00018****	-0.00061	-0.00012	-0.00012	-0.00061	**-0.00049****
	[-0.00034, -0.00002]	[-0.0014, 0.0002]	[-0.00029, 0.00004]	[-0.00029, 0.00004]	[-0.00144, 0.00021]	[-0.00090, -0.00008]
Post neonatal SUDI in first year	**-0.00104*****	-0.00162	**-0.00186****	**-0.00109****	-0.00044	-0.00012
	[-0.00171, -0.00036]	[-0.0064, 0.0032]	[-0.00366, -0.00005]	[-0.00209, -0.00009]	[-0.00132, 0.00043]	[-0.00076, 0.00051]

Note: Parameter estimates statistically different from zero at 99% (***), 95% (**), 90% (*) confidence.

Table D3 cont. Comparison of PSM results for various specifications (2009-2011 births)

	teffect nnmatch		psmatch2			
Comparison *Estimation* *Adjustment*	Outside FS TLAs ATE (1)	Outside FS TLAs ATT (2)	Outside FS TLAs ATT (3)	Outside FS TLAs ATT matching adj (4)	Inside FS TLAs ATT (5)	Inside FS TLAs ATT matching adj (6)
OT/CYF Report of Concern or other initial intake event in first year	**0.1609***** [0.0811, 0.2408]	**0.14725***** [0.1192, 0.1753]	**0.1423***** [0.1183, 0.1664]		**0.1719***** [0.1520, 0.1919]	
OT/CYF Report of Concern or other initial intake event in second year	**0.0748**** [0.0025, 0.1472]	**0.08850***** [0.0631, 0.1139]	**0.0852***** [0.0636, 0.1068]		**0.0918***** [0.0736, 0.1101]	
OT/CYF care placement in first year	0.0026 [-0.0128, 0.0180]	0.00440 [-0.0048, 0.0136]	0.0067 [-0.0016, 0.0149]		0.0048 [-0.0015, 0.0111]	
OT/CYF care placement in second year	**0.0252***** [0.0074, 0.0430]	**0.01920***** [0.0048, 0.0336]	**0.0182***** [0.0051, 0.0313]		**0.0197***** [0.0095, 0.0298]	
Hospitalised for a maltreatment related injury in first year	**0.0034*** [-0.0006, 0.0073]	0.00118 [-0.0040, 0.0064]	**0.0047**** [0.0011, 0.0083]		**0.0034**** [0.0002, 0.0065]	
Hospitalised for a maltreatment related injury in second year	0.0026 [-0.0032, 0.0083]	0.00078 [-0.0031, 0.0047]	0.0005 [-0.0028, 0.0038]		-0.0001 [-0.0032, 0.0029]	
Hospitalised for a long bone fracture in first year	0.0000 [-0.0024, 0.0024]	-0.00064 [-0.0049, 0.0036]	0.0017 [-0.0006, 0.0039]		0.0017 [-0.0003, 0.0037]	

Note: Parameter estimates statistically different from zero at 99% (***), 95% (**), 90% (*) confidence.

Table D3 cont. Comparison of PSM results for various specifications (2009-2011 births)

	teffect nnmatch		psmatch2			
Comparison *Estimation* *Adjustment*	Outside FS TLAs ATE (1)	Outside FS TLAs ATT (2)	Outside FS TLAs ATT (3)	Outside FS TLAs ATT matching adj (4)	Inside FS TLAs ATT (5)	Inside FS TLAs ATT matching adj (6)
Enrolled with a PHO in first year	-0.0005	-0.00557	-0.0088		**0.0168****	
	[-0.0083, 0.0073]	[-0.0187, 0.0076]	[-0.0265, 0.0088]		[0.0018, 0.0317]	
Enrolled with a PHO in second year	0.0025	-0.00572	-0.0054		0.0092	
	[-0.0015, 0.0065]	[-0.0154, 0.0040]	[-0.0218, 0.0109]		[-0.0043, 0.0226]	
Fully immunised at 1+ milestone age in first year	0.0252	**0.04594*****	**0.0401*****		**0.0536*****	
	[-0.0088, 0.0592]	[0.0230, 0.0689]	[0.0181, 0.0620]		[0.0355, 0.0716]	
Fully immunised at every milestone age in first year	**0.0478***	**0.05589*****	**0.0454*****		**0.0440*****	
	[-0.0002, 0.0957]	[0.0254, 0.0864]	[0.0193, 0.0716]		[0.0219, 0.0661]	
Fully immunised at 1+ milestone age in second year	0.0273	**0.02062****	0.0205*		**0.0437*****	
	[-0.0053, 0.0599]	[0.0025, 0.0387]	[-0.0004, 0.0415]		[0.0266, 0.0608]	
Fully immunised at every milestone age in second year	**0.0371***	**0.05020*****	**0.0302****		**0.0343*****	
	[-0.0041, 0.0782]	[0.0213, 0.0791]	[0.0040, 0.0564]		[0.0132, 0.0554]	
Enrolled with an ECE provider	-0.0504	-0.02662*	**-0.0279****		0.0141	
	[-0.1304, 0.0295]	[-0.0573, 0.0040]	[-0.0551, -0.0007]		[-0.0081, 0.0363]	
Attended B4SC	-0.0190	0.02588*	0.0172		0.0166	
	[-0.0877, 0.0497]	[-0.0023, 0.0541]	[-0.0080, 0.0424]		[-0.0038, 0.0371]	

Note: Parameter estimates statistically different from zero at 99% (***), 95% (**), 90% (*) confidence.

Impact of different PSM specifications on main results for the 2009-2015 birth cohorts

This section provides more information of how the changes we made to the PSM methods affected the results we obtained for our main analysis of 2009-2015 birth cohorts, in particular, our decision to select controls from TLAs where Family Start was available rather than where it was not available, and then to modifying our matching approach to better estimate impacts on mortality rates.

Table D4 below shows the *characteristics of the matched samples* when controls are selected from TLAs where Family Start was available (right-hand side panel), rather than where it was not available (left-hand side panel). Note that the number of participants matched to at least one control is materially higher (16,458 compared to 15,828, where the total number of participants is 16,764). The number of controls included is also significantly larger (45,474 compared to 26,151). Clearly there are more non-participant children who share the same observed characteristics as participants in TLAs where Family Start is available. An advantage of the larger sample size is that impacts are more precisely estimated.

In both cases the differences between matched participants and controls are quite small, although selecting controls from TLAs where Family Start was not available leads to larger differences. There is marked difference in the proportion who were born in a major urban area. This arises because Family Start is available in Auckland, where a significant proportion of the population is Pasifika. The exact matching criteria ensure that Pasifika participants are matched to Pasifika controls, but the combining of major urban location and Pasifika ethnicity in the exact matching leads to Pasifika participants in Auckland being matched to Pasifika participants outside of Auckland (or other Family Start TLA) who don't live in major urban area.

Table D5 shows the impacts on estimates when controls are selected from TLAs where Family Start was available, rather than where it was not available, and the impact of modifying the matching method for mortality outcomes.

Selecting controls from the same TLAs as participant children, rather than from TLA where Family Start is not available sees the estimated impact of Family Start on ECE participation change from a significant decrease of 1.6 per 100 children to an increase of 1.1 per 100 children (significant at the 10% level but not at the 5% level). The estimated impact on PHO enrolment rates in the first year increases from 0.85 to 1.6 per 100 children. The estimated impact on PHO enrolment rates in the second year increases from 0.34 to 0.70 per 100 children (and is now significant). The estimated impact on post-neonatal mortality is also greater, increasing from 2.1 per 1000 children (CI, 0.3 to 4.0) to 3.1 per 1000 children (CI, 1.8 to 4.3), deaths from SUDI and injury in the first year are largely unchanged, as are estimated impacts on immunisation rates which are positive and significant.

Implementing the modified matching method for mortality outcomes reduces the estimated impacts by half or more. The estimated impact on post-neonatal mortality in the first year is reduced from 3.1 per 1000 children (CI, 1.8 to 4.3) to 1.2 per 1000 children (CI, 0.2 to 2.1). The estimated impact on SUDI deaths reduces from 1.4 per 1000 to 0.66 per 1000 children. The estimated impact on injury deaths in the first year reduces from 0.91 per 1000 children to 0.56 per 1000 children. Although materially smaller, the three mortality impacts in the first year are significant at the 5% level.

The impact of the two changes together (i.e. comparing Column 1 with 4) reduces the estimated reduction in neo-natal mortality in the first year from 2.1 per 1000 children (CI, 0.3 to 4.0) to 1.2 per

1000 children (CI, 0.3 to 2.1), SUDI deaths from 1.5 per 1000 children (CI, -0.0 to 3.0) to 0.66 per 1000 children (CI, 0.06 to 1.25) and injury deaths in the first year from 1.3 per 1000 children (CI, -0.14 to 2.8) to 0.56 per 1000 children (CI, 0.13 to 1.0). However, the estimated impact on injury deaths is smaller but is more significant.

Table D4. Difference between mean characteristics of matched participants and comparisons, by PSM specification

	Comparisons drawn from TLAs without Family Start			Comparisons drawn from TLAs with Family Start		
	Matched participants	Matched comparisons	Difference	Matched participants	Matched comparisons	Difference
Number of participants	15,828	26,151		16,458	45,474	
Māori	0.6067	0.6067	0.0000	0.5982	0.5982	0.0000
Pasifika	0.2846	0.2930	-0.0085	0.2982	0.3087	-0.0105
Pasifika and not Māori	0.1730	0.1730	0.0000	0.1845	0.1845	0.0000
Female	0.4819	0.4778	0.0042	0.4813	0.4744	0.0070
Child had a low birth weight	0.0961	0.0892	0.0069	0.0968	0.0984	-0.0016
Child was born in NZDEP 9-10 area	0.6327	0.6327	0.0000	0.6460	0.6460	0.0000
Child was born in a major urban area	0.4885	0.3621	0.1264	0.5021	0.5076	-0.0055
Child is their mother's first born	0.4195	0.4386	-0.0191	0.4245	0.4333	-0.0088
Mother						
Single at the birth of the child	0.6630	0.6481	0.0149	0.6672	0.6721	-0.0050
Supported by benefit within 3 months of child's birth	0.7020	0.7020	0.0000	0.7000	0.7000	0.0000
Under 18 when child was born	0.0485	0.0447	0.0038	0.0571	0.0591	-0.0020
Under 20 when child was born	0.2003	0.2003	0.0000	0.2159	0.2159	0.0000
Under 25 when child was born	0.5519	0.5542	-0.0023	0.5606	0.5640	-0.0033
Smoked at the time of the child's birth	0.3809	0.3962	-0.0153	0.3827	0.3789	0.0038
Supported by benefits for 3 or more of the last 5 years	0.3166	0.2919	0.0247	0.3131	0.3064	0.0067
Received mental health services or prescription in the 5 years before the child's birth	0.3154	0.3241	-0.0087	0.3243	0.3194	0.0049
Had a child protection placement before age 18	0.1068	0.1015	0.0052	0.1136	0.1074	0.0062
Had a child protection notification before age 18	0.4471	0.4446	0.0025	0.4587	0.4542	0.0045
Sibling had a child protection placement during the 5 years prior to child's birth	0.0425	0.0347	0.0079	0.0469	0.0407	0.0062
Sibling had a child protection notification during the 5 years prior to child's birth	0.3369	0.3224	0.0144	0.3398	0.3352	0.0046
Served a community service sentence during the 5 years prior to child's birth	0.1555	0.1577	-0.0022	0.1576	0.1512	0.0064
Served a custodial sentence during the 5 years prior to child's birth	0.0306	0.0311	-0.0005	0.0328	0.0289	0.0039
Father						
Recorded on birth certificate and linked to IDI spine	0.8601	0.8744	-0.0142	0.8568	0.8537	0.0032

Received an incapacity benefit due to substance abuse in the 5 years prior to child's birth	0.0343	0.0406	-0.0063	0.0349	0.0338	0.0011
Received an incapacity benefit due to mental health in the 5 years prior to child's birth	0.0788	0.0977	-0.0189	0.0792	0.0799	-0.0008
Had a child protection placement before age 18	0.0812	0.0790	0.0022	0.0836	0.0818	0.0018
Had a child protection notification before age 18	0.2974	0.3123	-0.0149	0.3002	0.2976	0.0026
Served a community service sentence during the 5 years prior to child's birth	0.3234	0.3379	-0.0146	0.3232	0.3205	0.0027
Served a custodial sentence during the 5 years prior to child's birth	0.1487	0.1555	-0.0068	0.1503	0.1447	0.0056

Notes: Shared are based on randomly rounded counts with accordance to Stats NZ's random rounding 3 regulations. For full list of variables, see Appendix C. Siblings refers to all other children linked to the child's mother using DIA data; this can include birth siblings who no longer live in the same household as the mother or participant child, and exclude step-siblings or other unrelated children who live in the same household as the mother or participant child.

Table D5. Comparison of PSM results for various specifications (2009-2015 births)

	Estimate (95% CI)			
Comparison	Outside FS TLAs (A)	Outside FS TLAs matching adj (B)	Inside FS TLAs (C)	Inside FS TLAs matching adj (D)
Number of participants	14,511	14,511	14,511	14,511
Number of matched participants	13,710	13,710	14,244	14,238
Number of matched comparisons	22,602	22,440	40,926	40,473
Weighted and matched comparisons	13,710	13,710	14,244	14,238
Post neonatal mortality in first year	**-0.00214**** [-0.00397, -0.00030]	-0.00040 [-0.00145, 0.00065]	**-0.00305***** [-0.00430, -0.00180]	**-0.00120**** [-0.00212, -0.00029]
Post neonatal mortality in second year	-0.00049 [-0.00136, 0.00039]	-0.00045* [-0.00097, 0.00007]	-0.00051 [-0.00111, 0.00010]	-0.00035 [-0.00083, 0.00013]
Post neonatal mortality in third to sixth years	0.00048 [-0.00052, 0.00148]	0.00060 [-0.00032, 0.00153]	0.00020 [-0.00091, 0.00130]	0.00024 [-0.00086, 0.00133]
Post neonatal injury death in first year	-0.00133* [-0.00279, 0.00014]	-0.00057* [-0.00122, 0.00009]	**-0.00091***** [-0.00158, -0.00024]	**-0.00056**** [-0.00100, -0.00013]
Post neonatal injury death in second year	-0.00002 [-0.00053, 0.00049]	-0.00002 [-0.00036, 0.00032]	-0.00010 [-0.00048, 0.00028]	-0.00002 [-0.00033, 0.00030]
Post neonatal injury death in third to sixth years	0.00041 [-0.00015, 0.00097]	0.00041 [-0.00014, 0.00096]	0.00031 [-0.00025, 0.00087]	0.00031 [-0.00033, 0.00096]
Post neonatal SUDI in first year	-0.00147* [-0.00301, 0.00008]	-0.00056 [-0.00134, 0.00022]	**-0.00142***** [-0.00224, -0.00061]	**-0.00066**** [-0.00125, -0.00006]

Notes: Mortality outcomes in the first year are not observed for the 2015 cohort, second year outcomes are not observed for the 2014 & 2015 cohort, and sixth year outcomes are not observed for the 2011-2015 cohorts. Counts are based on rounded counts using Stats NZ's random rounding 3 regulations. For full list of variables, see Appendix C.

Table D5 cont. Comparison of PSM results for various specifications (2009-2015 births)

	Estimate (95% CI)	
Comparison	Outside FS TLAs (A)	Inside FS TLAs (C)
Number of participants	16,764	16,764
Number of matched participants	15,825	16,458
Number of matched comparisons	26,154	47,238
Weighted and matched comparisons	15,825	16,458
OT/CYF Report of Concern or other initial intake event in first year	**0.1275*** **	**0.1659*** **
	[0.1165, 0.1385]	[0.1572, 0.1747]
OT/CYF Report of Concern or other initial intake event in second year	**0.0772*** **	**0.0988*** **
	[0.0669, 0.0875]	[0.0907, 0.1069]
OT/CYF Report of Concern or other initial intake event in third to sixth year	**0.0268*** **	**0.0434*** **
	[0.0158, 0.0377]	[0.0350, 0.0518]
OT/CYF assessment in first year	**0.1079*** **	**0.1347*** **
	[0.0978, 0.1181]	[0.1266, 0.1429]
OT/CYF assessment in second year	**0.0584*** **	**0.0720*** **
	[0.0496, 0.0673]	[0.0650, 0.0791]
OT/CYF assessment in third to sixth year	**0.0160*** **	**0.0237*** **
	[0.0075, 0.0244]	[0.0172, 0.0303]
OT/CYF care placement in first year	**0.0071*** **	**0.0077*** **
	[0.0030, 0.0111]	[0.0044, 0.0110]
OT/CYF care placement in second year	**0.0147*** **	**0.0203*** **
	[0.0092, 0.0203]	[0.0159, 0.0247]
OT/CYF care placement in sixth year	0.0035*	**0.0063*** **
	[-0.0003, 0.0073]	[0.0033, 0.0092]
Hospitalised for a maltreatment injury in first year	**0.0033*** **	**0.0022*** **
	[0.0015, 0.0051]	[0.0007, 0.0036]
Hospitalised for a maltreatment injury in second year	0.0005	**0.0021*** **
	[-0.0015, 0.0024]	[0.0006, 0.0036]
Hospitalised for a maltreatment injury in sixth year	0.0020	-0.0006
	[-0.0004, 0.0044]	[-0.0026, 0.0014]
Hospitalised for long bone fracture in first year	0.0012*	**0.0010** **
	[0.0000, 0.0025]	[0.0001, 0.0019]
Enrolled with a PHO in first year	**0.0085** **	**0.0162*** **
	[0.0015, 0.0154]	[0.0117, 0.0207]
Enrolled with a PHO in second year	0.0034	**0.0070*** **
	[-0.0029, 0.0098]	[0.0032, 0.0108]
Enrolled with a PHO in sixth year	0.0022	0.0033*
	[-0.0046, 0.0090]	[-0.0005, 0.0072]
Fully immunised at 1+ milestone age in first year	**0.0392*** **	**0.0400*** **
	[0.0301, 0.0483]	[0.0338, 0.0463]
Fully immunised at every milestone age in first year	**0.0410*** **	**0.0367*** **
	[0.0287, 0.0534]	[0.0275, 0.0459]
Fully immunised at 1+ milestone age in second year	**0.0320*** **	**0.0296*** **
	[0.0230, 0.0411]	[0.0235, 0.0357]
Fully immunised at every milestone age in second year	**0.0375*** **	**0.0354*** **
	[0.0261, 0.0490]	[0.0269, 0.0438]
Fully immunised at every milestone by age six	0.0122*	**0.0307*** **

	[-0.0013, 0.0258]	[0.0210, 0.0405]
Enrolled with an ECE provider	-0.0164***	0.0105*
	[-0.0310, -0.0018]	[-0.0005, 0.0216]
Attended Before School Check (B4SC)	0.0315***	0.0219***
	[0.0186, 0.0445]	[0.0125, 0.0313]

Notes: Mortality outcomes in the first year are not observed for the 2015 cohort, second year outcomes are not observed for the 2014 & 2015 cohort, and sixth year outcomes are not observed for the 2011-2015 cohorts. Counts are based on rounded counts using Stats NZ's random rounding 3 regulations.

Additional tables – main PSM analysis

Table D6. Family Start participant mean outcomes (2009-2015 births), main PSM analysis-total

Year	First	Second	Sixth
		Total	
Post neonatal mortality	0.00169	0.00050	0.00098
Post neonatal injury death	0.00028	0.00025	0.00039
Post neonatal SUDI	0.00063		
OT/CYF Report of Concern or other initial intake event	0.3726	0.2625	0.1767
OT/CYF assessment	0.2848	0.1794	0.0993
OT/CYF care placement	0.0283	0.0584	0.0178
Child or sibling appeared in a Police FV report to OT	0.2814	0.2682	0.2011
Hospitalised for a maltreatment related injury	0.0063	0.0066	0.0079
Hospitalised for a long bone fracture	0.0026		
Enrolled with a PHO	0.9751	0.9896	0.9973
Fully immunised at 1+ milestone age	0.9178	0.9235	
Fully immunised at every milestone age	0.6430	0.7828	0.8208
Attended B4SC			0.8072
Significant issues identified in B4SC			0.2692
Other issues identified in B4SC			0.2567
Within a healthy weight range at B4SC			0.1243
Enrolled with an ECE provider			0.6746
Mother received publicly funded mental health services	0.2612	0.2421	0.1831
		Māori	
Post neonatal mortality	0.00223	0.00084	0.00100
Post neonatal injury death	0.00047	0.00042	0.00066
Post neonatal SUDI	0.00071		
OT/CYF Report of Concern or other initial intake event	0.4194	0.2999	0.2028
OT/CYF assessment	0.3219	0.2100	0.1139
OT/CYF care placement	0.0331	0.0711	0.0218
Child or sibling appeared in a Police FV report to OT	0.3305	0.3200	0.2476
Hospitalised for a maltreatment related injury	0.0070	0.0076	0.0081
Hospitalised for a long bone fracture	0.0028		
Enrolled with a PHO	0.9763	0.9906	0.9983
Fully immunised at 12 month milestone age	0.8996	0.9153	
Fully immunised at every milestone age	0.5733	0.7447	0.8106
Attended B4SC			0.7921
Significant issues identified in B4SC			0.2867
Other issues identified in B4SC			0.2598
Within a healthy weight range at B4SC			0.1236

Enrolled with an ECE provider			0.6603
Mother received publicly funded mental health services	0.2486	0.2317	0.1863
		Pasifika	
Post neonatal mortality	0.00118	0.00028	0.00066
Post neonatal injury death	0.00024	0.00000	0.00000
Post neonatal SUDI	0.00047		
OT/CYF Report of Concern or other initial intake event	0.3049	0.2027	0.1264
OT/CYF assessment	0.2278	0.1417	0.0766
OT/CYF care placement	0.0148	0.0353	0.0115
Child or sibling appeared in a Police FV report to OT	0.2608	0.2288	0.1605
Hospitalised for a maltreatment related injury	0.0055	0.0062	0.0073
Hospitalised for a long bone fracture	0.0025		
Enrolled with a PHO	0.9674	0.9867	0.9969
Fully immunised at 12 month milestone age	0.9278	0.9220	
Fully immunised at every milestone age	0.6527	0.7735	0.7846
Attended B4SC			0.7627
Significant issues identified in B4SC			0.2420
Other issues identified in B4SC			0.2219
Within a healthy weight range at B4SC			0.1738
Enrolled with an ECE provider			0.6323
Mother received publicly funded mental health services	0.1579	0.1397	0.0945

Notes: Shares are based on rounded counts using Stats NZ's random rounding 3 regulations. For full list of variables, see Appendix C.

APPENDIX E: ADDITIONAL DIFFERENCE IN DIFFERENCE (DID) ANALYSES

Comparison of study methodology with Vaithianathan et al. (2016)

As with the PSM methodology, one of the key tasks for this evaluation was to explore the extent to which the results reported in Vaithianathan et al. (2016) could be replicated. For this reason, our starting point in the development of our DiD approach was to replicate the previous evaluation's approach as much as possible. As with the PSM analyses, we also systematically explored potential improvements to that model, reflecting input from our technical advisor and EAG members, and the additional data available in the IDI.

Approach to DiD taken in Vaithianathan et al. (2016)

Vaithianathan et al. (2016) explored the area/TLA/community-level impact of the availability of the programme on all children who were supported by a benefit in the first 13 weeks after birth. This (child-benefit) inclusion criterion was selected to narrow analysis from all births within the TLA to a sub-group that was more likely to represent the target population of Family Start. TLA-year-quarter of birth was the unit of analysis, with outcomes aggregated (averaged) to that level (e.g., TLA 7, Q4 2006). Birth cohort and TLA fixed effects, and the proportion of births to mothers aged under 18 in each TLA-year-quarter were used as controls.

The following equation was then estimated for each outcome of interest (Vaithianathan et al., 2016, p 29):

$$Y_{kt} = \gamma FS_{kt} + \tau TEEN_{kt} + \alpha_k + \delta_t + u_{kt}$$

In the above equation, Y_{kt} represents the average of outcome Y across all births in TLA k , born in year-quarter t. α_k and δ_t are a set of birth-TLA and birth year-quarter dummies. γ is the coefficient of interest, capturing the average change in outcomes following the introduction of Family Start in treatment TLAs. FS was a dummy variable indicating whether Family Start was available in TLA k during year-quarter period t, while τ is the coefficient of the variable $TEEN_{kt}$ (share of proportion of teen births by TLA and year-quarter).

Vaithianathan et al. (2016) examined the 2005 – 2007 expansion and included data from children born from Q3 2004 – Q4 2011. As a result, the sample was limited to TLAs that received Family Start between 2005 and 2007 (treatment) and TLAs that did not receive the service during the study period (control).[109] Note that a number of TLAs have been removed (termed Semi-treated below) since the authors could not determine their treatment status.

[109] Allocation of children into control and treatment groups was based on mother address at birth (sourced from the Ministries of Health and Social Development records).

We built on the original approach by testing the following model settings:

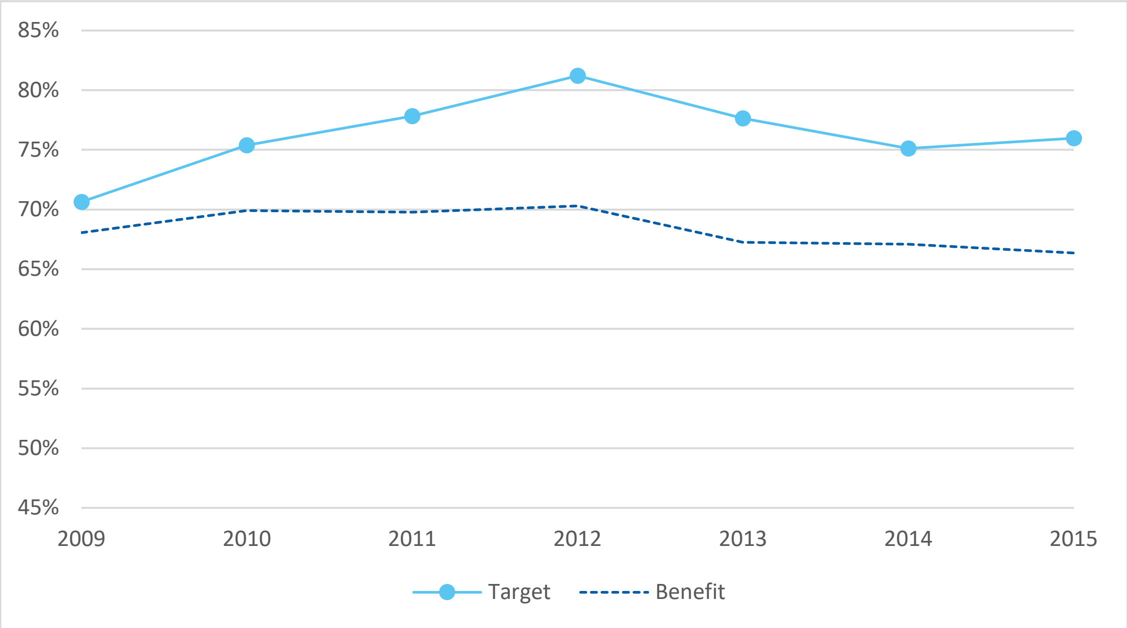
- whether the unit of analysis should be TLA year-quarter or whether analysis should be individual-level
- whether weights should be introduced to the model to adjust for differences in the numbers of relevant births in each TLA (when analysis is aggregated at the TLA year-quarter level)[110]
- whether including 'semi-treated' TLAs (where Family Start availability was uncertain) would impact model outcomes
- the extent to which the introduction of additional control variables could strengthen models.

Inclusion criterion / target group

As mentioned in the methodology section (in main text), we used a different inclusion criterion than Vaithianathan et al. (2016). This sub-section outlines the motivation for this decision. Figure E1 compares the share of Family Start participants (born between 2009 and 2015) that also satisfied our (alternative) inclusion criterion (termed 'target' thereafter), and the child-benefit based criterion used by Vaithianathan et al. (2016). The figure shows that in all years, our criterion included a greater share of actual Family Start participants. Over the entire period, 76% of Family Start participants satisfied our inclusion criterion, compared with 68% for the child-benefit criterion.

In addition, Family Start participants make up a greater share of the 'target' group than of the 'child-benefit' group (Figure E2 below). This was an important factor in our decision to adopt the 'target' criterion for our main DiD analyses, as it increased the likelihood of detecting the effects of the programme (i.e., increasing the statistical power).

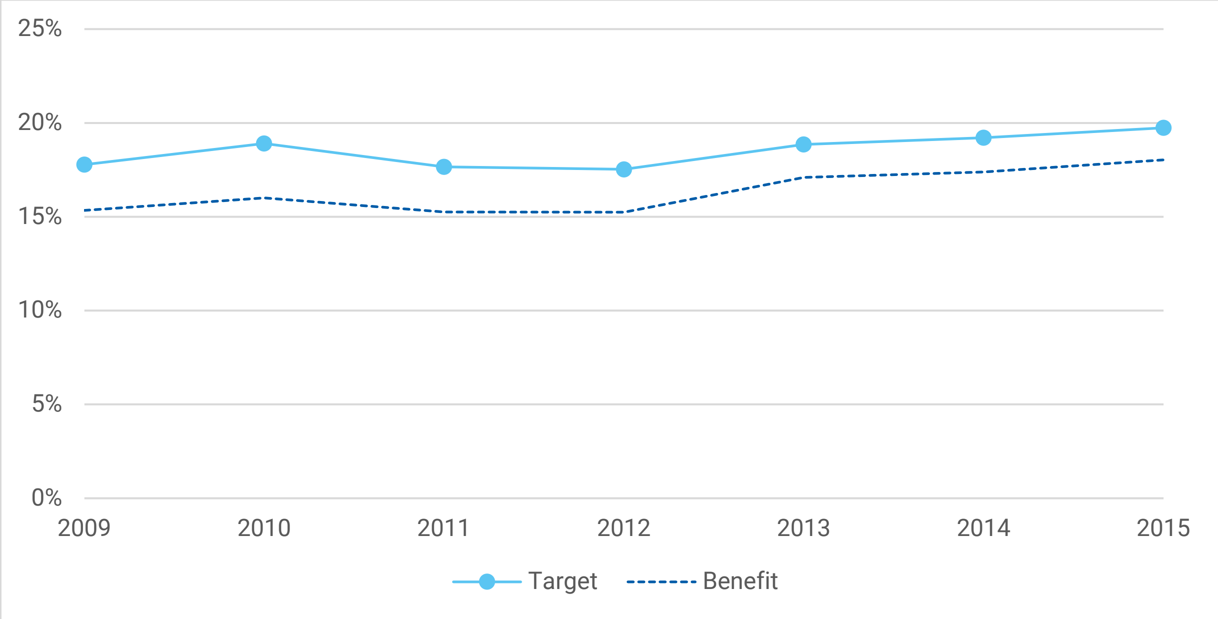
Figure E1. Proportion of Family Start participants meeting each inclusion criterion



Notes: Shares are based on rounded counts with accordance to Stats NZ's random rounding 3 regulations. For full list of variables, see Appendix C.

110 This results in TLAs with a greater number of births having importance when calculating the average effect of the programme. For example, the average change is less sensitive to changes that occurred in TLAs with small number of births TLAs (e.g. Buller District), compared to large ones (e.g. Manukau Distinct).

Figure E2. Proportion of children meeting a given criterion who participated in Family Start



Notes: Shares are based on rounded counts with accordance to Stats NZ's random rounding 3 regulations. For full list of variables, see Appendix C.

Model comparisons

Table E1 presents the number of TLAs, and the size of target, treatment, and control groups in Vaithianathan et al. (2016) and our replica sample. This table is analogous to Table 2 from Vaithianathan et al. (2016). The table shows that Vaithianathan et al. (2016) had two additional control TLAs. These are the Banks Peninsula District and the Territorial Authority residual category (which was not used in Vaithianathan et al. [2016]). In our study, we use a more recent TLA classification (2010), in which the Banks Peninsula District is included within Christchurch City. As discussed in the main body, we do not except this to significantly affect the results due to low birth count. Overall, we find that our replication sample is smaller, with our replica treatment group about 6 percentage points smaller than in Vaithianathan et al. (2016), and the control group by 8 percentage points. In addition, the proportion of children that satisfy the benefit criterion from the overall child population is slightly smaller in our replication sample.

Table E1. Number of births by treatment status of TLA, for births between Q3 2004 and Q4 2011

	Replication			Vaithianathan et al. (2016)			Meeting criterion (share)	
	Number of TLAs	Births	Target population	Number of TLAs	Births	Target population	Replication	Vaithianathan et al. (2016)
Control	35	140,841	23,082	37	145,581	25,110	0.1639	0.1725
Treatment	14	140,649	35,283	14	142,323	37,611	0.2509	0.2643
Phase-in TLAs (Treatment group) by quarter service commenced								
2005Q3	2	49,485	8,418	2	50,037	9,066	0.1701	0.1812
2006Q1	2	60,351	17,343	2	61,032	18,450	0.2874	0.3023
2006Q3	4	21,798	6,828	4	22,074	7,227	0.3132	0.3274
2006Q4	4	4,614	1,269	4	4,788	1,371	0.2750	0.2863
2007Q4	2	4,401	1,425	2	4,389	1,497	0.3238	0.3411

Notes: Counts and shares have been randomly rounded in accordance with Stats NZ's random rounding 3 regulations. For a full list of variables, see Appendix C.

Table E2 presents the percentage of children who participated in Family Start amongst those born in treatment ('phased-in') TLAs who satisfied the benefit inclusion criteria (an equivalent analysis to that presented in Table 3 in Vaithianathan et al. 2016).The previous study found that about 15% of all children linked to main benefits via their mothers in the first 13 weeks of life were enrolled to Family Start (and 5.8% of all births). Our replication results show a similar overall share, and slightly (i.e., under 1 percentage point) lower shares for Māori and Pasifika children. Finally, shares of total births in phase-in TLAs were very similar in both studies.

Table E2. Estimated programme coverage in phase-in TLAs, for children born 2009-2011, by sub-group

	Family Start participation rates among children who met the benefit inclusion criteria			Total births in phase-in TLAs
	All	Māori	Pasifika	
Replication	0.150	0.174	0.143	0.059
Vaithianathan et al. (2016)	0.151	0.181	0.149	0.058

Notes: Shares are based on rounded counts in accordance with Stats NZ's random rounding 3 regulations. For a full list of variables, see Appendix C.

Table E3 presents the proportion of children from the control TLAs recording year one outcomes. The rightmost column shows the results from Table 4B in Vaithianathan et al. (2016), and our replication in middle column. Note that due to data quality and coverage issues, we did not replicate (here or elsewhere) a number of outcomes used in Vaithianathan et al. (2016).[111] Overall, the table shows that mortality outcomes and long-bone fracture hospitalisation are less frequent in our replication. On the other hand, other outcomes are more common in our replication (differences varying from 0.07 to 1.1 percentage points, leading to large percentage differences).

Table E3. Means of first year outcomes in the control group, Q3 2004 - Q4 2011

Outcome	Replication (*N*= 23,079)	Vaithianathan et al. (2016) (*N*= 25,110)
Post-neonatal mortality	0.0026	0.0034
Post-neonatal injury death	0.0006	0.0010
Post-neonatal SUDI	0.0013	0.0017
OT/CYF Report of Concern or other initial intake event	0.210	0.199
OT/CYF care placement	0.015	0.014
Enrolled with PHO	0.970	0.961
Hospitalised for a long bone fracture	0.001	0.002
Hospitalised for a maltreatment related injury	0.003	0.001

Notes: Shares are based on rounded counts in accordance with Stats NZ's random rounding 3 regulations. For a full list of variables, see Appendix C.

Table E4 presents the DiD (Area Level) estimates from Vaithianathan et al. (2016), followed by our best attempt to replicate the results, and when introducing variations to the model. Column 1 shows the results from Vaithianathan et al. (2016), with our replication results in Column 2. In terms of magnitude, our estimates are largely comparable, where estimates are within the

[111] For example, PrimHD data (used for substance abuse/mental health) is only systematically collected in the IDI from late 2007. In addition, we did not replicate finding of abuse due to data quality issues. Finally, our measure of maltreatment includes forms of diagnosis not included in Vaithianathan et al. (2016), leading to greater shares.

confidence intervals of those from Vaithianathan et al. (2016) (i.e., not statistically different). The key difference between Column 1 and Column 2 is that we are not able to replicate a statistically significant reduction in post-neonatal mortality (and the point estimate is smaller).

Column 3 estimates the model with population weights (based on the relative size of each TLA and birth-cohort). With weights, the estimate of post-neonatal mortality is now statistically significant. In addition, estimates with weights also find a statistically significant reduction in post-neonatal injury death (which was not found in the previous study). Column 4 estimates the model when a number of additional TLAs are included in the sample (i.e., semi-treated TLAs), and Column 5 includes a set of controls. Both types of modification do not materially change the estimates in Column 2. For Column 5, this is somewhat expected since Vaithianathan et al. (2016) reported that including controls in their model did not materially change their results. Column 6 estimates the model using weights, the additional TLAs, and the full set of controls. Here we estimate a statistically significant reduction in post-neonatal mortality, and a nearly statistically significant (at the $p < 0.1$ threshold) reduction in post-neonatal injury related death. Column 7 repeats the specification in Column 6, using child level observations (rather than aggregating observations to the TLA level).[112] Most outcomes do not show a significant difference between Columns 6 and 7. However, now post-neonatal injury related death is statistically significant (at the $p<0.05$ threshold).

Finally, we re-estimated all specifications using 2003q1 as the start period (rather than 2004q3) and found that the two mortality outcomes were no longer statistically significant (for all specifications). It seems that mortality rates in the control TLAs were lower in 2004q3-2005q2 compared with earlier and later periods, where a similar difference was not observed for the treatment TLAs. In addition, we re-examined our main specification, finding statistically significant reductions in overall post-neonatal mortality (and post-neonatal injury-related deaths) if the start period is changed (from 2003q1) to 2004q3. It is clear that the statistical significance of these mortality outcomes is sensitive to the period examined.

[112] We do not include weights in this specification; the effect of TLA relative size affects the point estimate by construction.

Table E4. Comparison of DiD results, Q3 2003 - Q4 2011

	V.(2016) original results for comparison	V.(2016) model replication using IDI data	Inclusion of weights	Inclusion of semi-treated TLAs	Inclusion of controls	Inclusion of semi-treated TLAs, weights, and controls	Individual level with semi-treated TLAs and controls
Outcome	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Post neonatal mortality	**-0.0035****	-0.002197	**-0.002596****	-0.002229	-0.001803	**-0.002606****	**-0.002612*****
	[-0.0067, -0.0003]	[-0.005002 - 0.000608]	[-0.004630 - -0.000563]	[-0.004952 - 0.000494]	[-0.004818 - 0.001211]	[-0.004770 - -0.000443]	[-0.004528 - -0.000696]
Post neonatal SUDI	-0.0006	0.000014	-0.000566	-0.000044	0.000414	-0.000357	-0.000548
	[-0.0022, 0.0010]	[-0.001531 - 0.001560]	[-0.002068 - 0.000936]	[-0.001497 - 0.001409]	[-0.001115 - 0.001942]	[-0.001919 - 0.001206]	[-0.001934 - 0.000838]
Post neonatal injury death	-0.0006	-0.000423	**-0.000806****	-0.000449	-0.000080	**-0.000683***	**-0.000774****
	[-0.0016, 0.0004]	[-0.001634 - 0.000788]	[-0.001566 - -0.000045]	[-0.001574 - 0.000676]	[-0.001252 - 0.001091]	[-0.001420 - 0.000054]	[-0.001512 - -0.000036]
OT/CYF Report of Concern or other initial intake event	0.0062	0.022475	0.007150	0.021420	**0.030527***	0.012672	0.005744
	[-0.028, 0.041]	[-0.004487 - 0.049438]	[-0.009590 - 0.023891]	[-0.004468 - 0.047307]	[-0.000566 - 0.061620]	[-0.004379 - 0.029723]	[-0.010078 - 0.021565]
OT/CYF care placement commenced	0.0042	0.010667	0.000694	0.010217	**0.010123***	0.001269	0.001611
	[-0.004. 0.013]	[-0.003720 - 0.025055]	[-0.003162 - 0.004549]	[-0.003929 - 0.024363]	[-0.001324 - 0.021570]	[-0.002705 - 0.005243]	[-0.002152 - 0.005374]
Enrolled with a PHO	-0.0042	0.005346	-0.005719	0.006345	0.003847	-0.005959	-0.004188
	[-0.024, 0.015]	[-0.010550 - 0.021241]	[-0.021705 - 0.010268]	[-0.008688 - 0.021377]	[-0.010633 - 0.018328]	[-0.019165 - 0.007247]	[-0.019496 - 0.011120]
Hospitalised for a long bone fracture	-0.0006	0.000003	0.000018	0.000190	0.000364	0.000228	-0.000009
	[-0.002, 0.001]	[-0.001918 - 0.001923]	[-0.000913 - 0.000949]	[-0.001648 - 0.002028]	[-0.002951 - 0.003679]	[-0.000934 - 0.001391]	[-0.000912 - 0.000894]
Hospitalised for a maltreatment related injury	0.0012	0.000997	-0.000452	0.001100	0.001637	0.000084	-0.000462
	[-0.002, 0.005]	[-0.001642 - 0.003636]	[-0.002020 - 0.001116]	[-0.001338 - 0.003538]	[-0.002388 - 0.005662]	[-0.001834 - 0.002002]	[-0.002006 - 0.001082]
Unit of analysis	TLA	TLA	TLA	TLA	TLA	TLA	Child
Semi-treated TALs included	No	No	No	Yes	No	Yes	Yes
Weighted to population size	No	No	Yes	No	No	Yes	-
Controls included	No	No	No	No	Yes	Yes	Yes

Note: Parameter estimates statistically different from zero at 99% (***), 95% (**), 90% (*) confidence.

Parallel trends tests

One of the key underlying assumptions of the Difference-in-Difference (DiD) methodology is the parallel trends assumption. This assumption requires outcomes in the treatment and control groups to follow the same (parallel) trend in the pre-treatment period, or to remain constant. In the context of this evaluation, outcomes of the treatment ('phase-in') and control ('never treated') TLAs are required to follow the same trend before the introduction of Family Start into 'phase-in' TLAs (between 2005q2 and 2007q3).

The test for parallel trends can be written as:

$$Y_{ikt} = \beta D_{kt} + \lambda(\delta_t * Treated_k) + \gamma X_{ikt} + \delta_k + \delta_t + \varepsilon_{ikt}$$

Where Y represents outcome for child *i,* born in TLA *k* in year-quarter *t*. D_{kt} indicates whether Family Start was available birth-TLA for children born in year-quarter *t* (same meaning as used in the main analysis). The interaction term between $Treated_k$ and δ_t captures whether the children from each *t* year-quarter cohorts were born in a treatment ('phase-in') TLA, regardless of whether the programme was available. As in the main analysis, X_{ikt} is a set of child-related controls, while δ_k and δ_t are birth-TLA and birth year-quarter fixed effects. In this analysis, the parameter of interest is λ, which captures the difference in average outcomes for children born in treated TLAs in each year-quarter prior to the programme becoming available were different from those experienced by children born in control TLAs (captured by δ_j). The use of dummy variables (rather than a linear time-trend variable) allows us to capture non-linearities in this trend.

We present the results of the tests graphically in Figure E3-Figure E9. which present differences in pre-treatment outcomes (between the control and treatment groups), for each λ. The mean outcomes of the treatment group are represented relative to the trend in outcomes recorded for the control group. The second quarter of 2007 is set to zero (reference group; horizontal line) by construction. Note that deviations (λ) are the residual in outcome when using the entire set of controls, and both the point estimates, and 95% confidence intervals are presented.

For the parallel trends assumption to hold, all λ should all be equal. To examine this more formally, we conducted a series of *F*-tests for joint significant of all λ (Table E5). The test null hypothesis for this test is rejected if the test score (*F*-score) is equivalent to a *p*-value of 5% (0.05) or lower. For brevity, we only present the results for the outcomes that were found to be statistically significant in the main DiD analyses.[113]

Test results

The results for the mental health service use presented in Figure E3 (first year) and Figure E4 (second year) across all children show similar patterns. In both figures, the treatment group shows a greater share of mothers using mental health services in early quarters, followed by a reduction to a level below that of the control group in later quarters. In most quarters, the differences are not statistically different than those of the control group, and from one another. However, the joint

[113] For all children, these were decrease in mother mental health services (in year 1 and 2), and an increase (decrease) in Family Violence contact-records/notifications in the first (sixth) year. For Māori, these were increases in notifications (initial assessment phase) and Family Violence contact-records/notifications in the first year. For Pasifika children, this was a reduction in Family Violence contact-records/notifications in the sixth year.

significance tests conclude that the null hypothesis is rejected, hence the parallel trend assumption is violated.

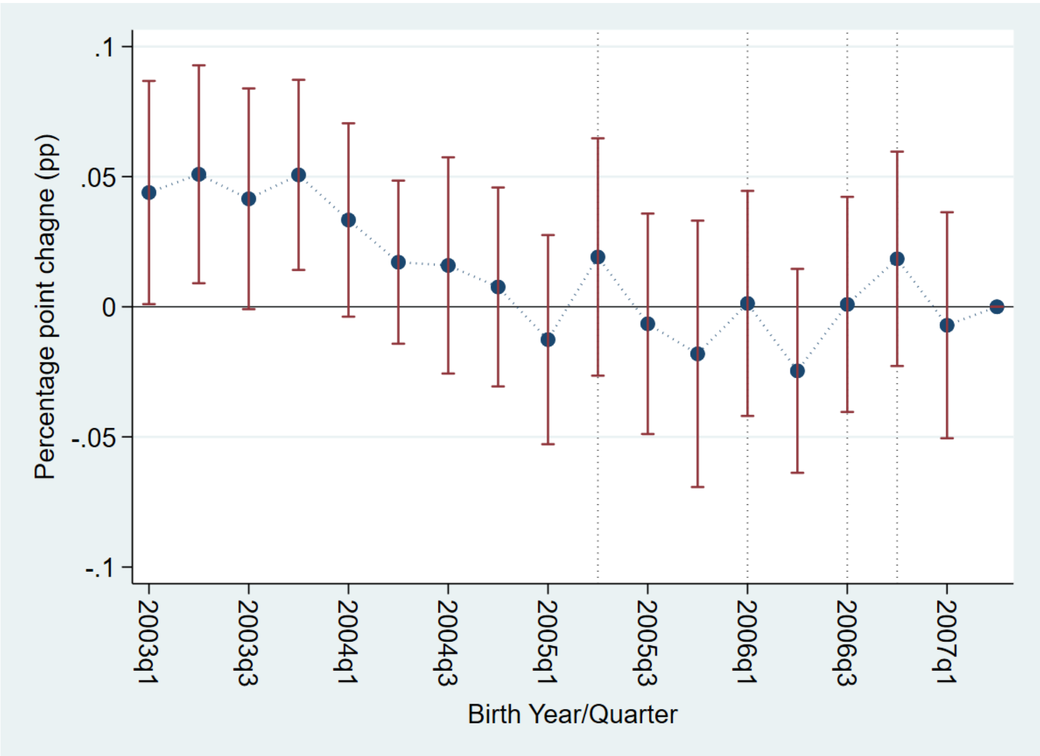
Figure E5 and

Figure E6 present the test results for family violence notifications in the first and sixth year for all children. For the first-year outcome, the figure suggests that the share of family violence events was lower in earlier quarters for the treatment group and converged to those of the control group over time. For the first year, the joint-significant test cannot be rejected, suggesting that the parallel trends assumption holds. For the sixth year, the figure suggests a decline in reports for the treatment group, with the joint significance test suggests that the parallel trend assumption is rejected.

Figure E7 and Figure E8 present the results for Māori. Neither figure presents a clear pre-existing trend and they do not seem to be statistically different. However, the figures show that the estimates deviate below and above the mean outcomes of the control group. The hypothesis for the test results for the Family Violence Notification cannot be rejected at the 5% level (only at the 10% level), while the results for the any Oranga Tamariki notification is rejected at the 1% level.

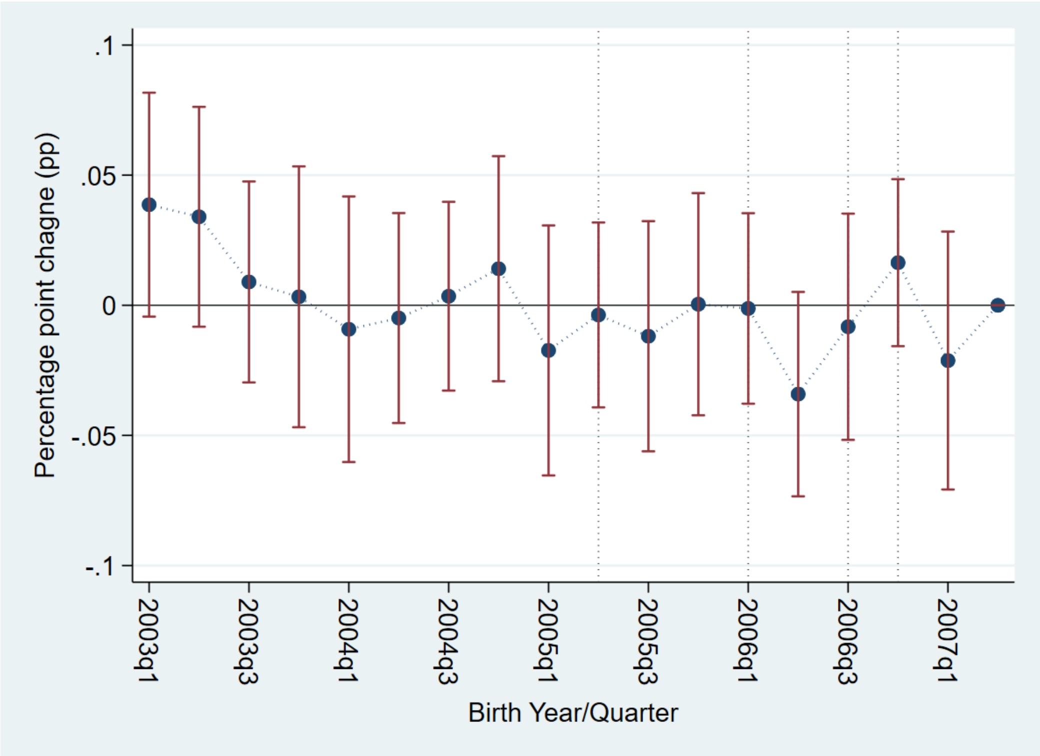
Finally, family violence events in the sixth year for Pasifika children is presented in Figure E9. There does not seem to be a clear pre-pattern, though difference in outcome deviate above/below the mean level of the control group. The results from the parallel trends test suggest that the parallel trends assumption is rejected at the 5% level.

Figure E3. Parallel trends test Mental Health Service use, year 1 (all ethnicities)



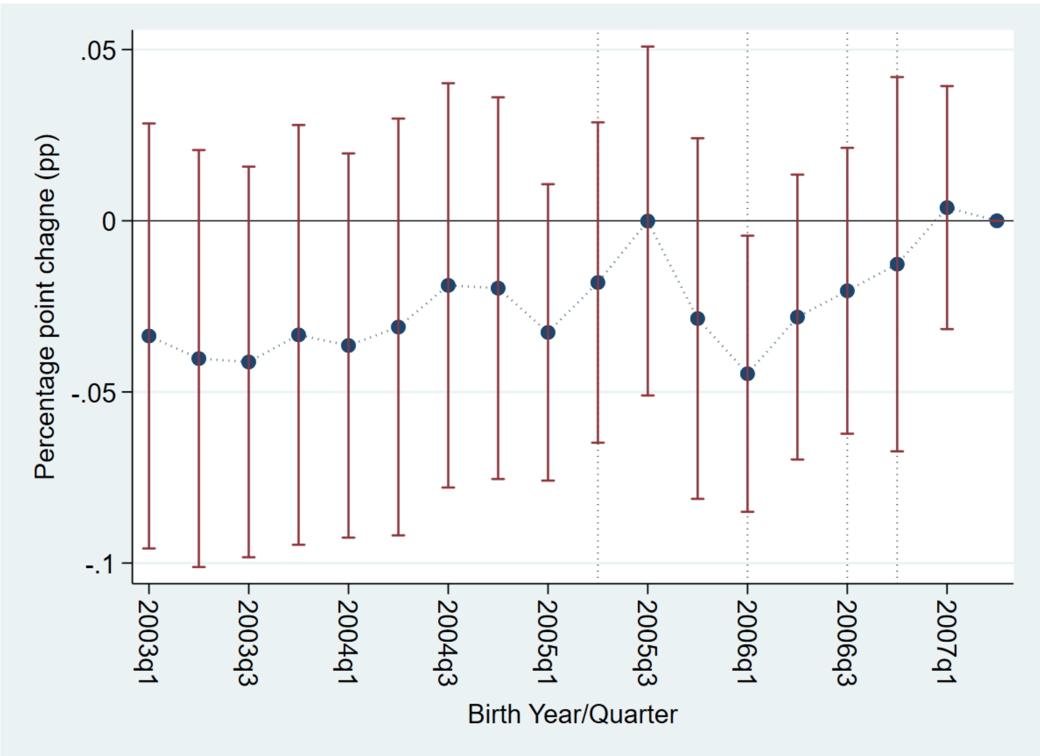
Notes: This figure shows the difference in mean outcomes of the treatment group for each birth year-quarter. Q2 2007 is set as a benchmark and equal to zero by construction. The vertical dotted lines show the year-quarter in which Family Start expanded into a new set of treatment TLAs.

Figure E4. Parallel trends test Mental Health Service use, year 2 (all ethnicities)



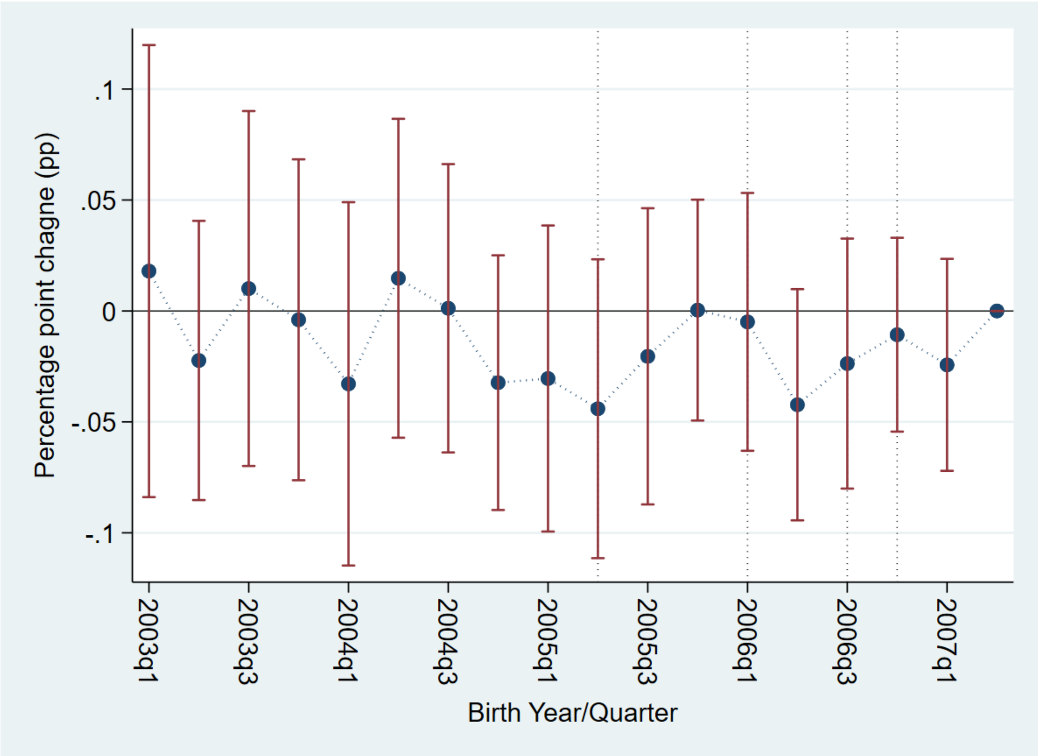
Notes: This figure shows the difference in mean outcomes of the treatment group for each birth year-quarter. Q2 2007 is set as a benchmark and equal to zero by construction. The vertical dotted lines show the year-quarter in which Family Start expanded into a new set of treatment TLAs.

Figure E5. Parallel trends test Family Violence event, year 1 (all ethnicities)



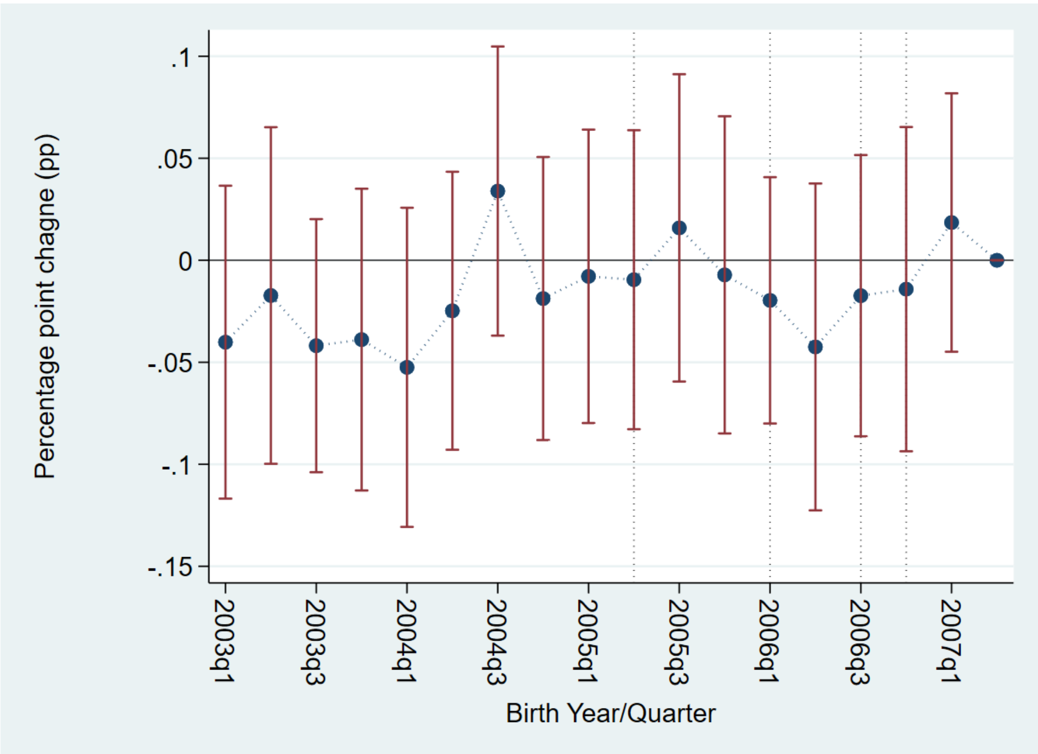
Notes: This figure shows the difference in mean outcomes of the treatment group for each birth year-quarter. Q2 2007 is set as a benchmark and equal to zero by construction. The vertical dotted lines show the year-quarter in which Family Start expanded into a new set of treatment TLAs.

Figure E6. Parallel trends test Family Violence event, year 6 (all ethnicities)



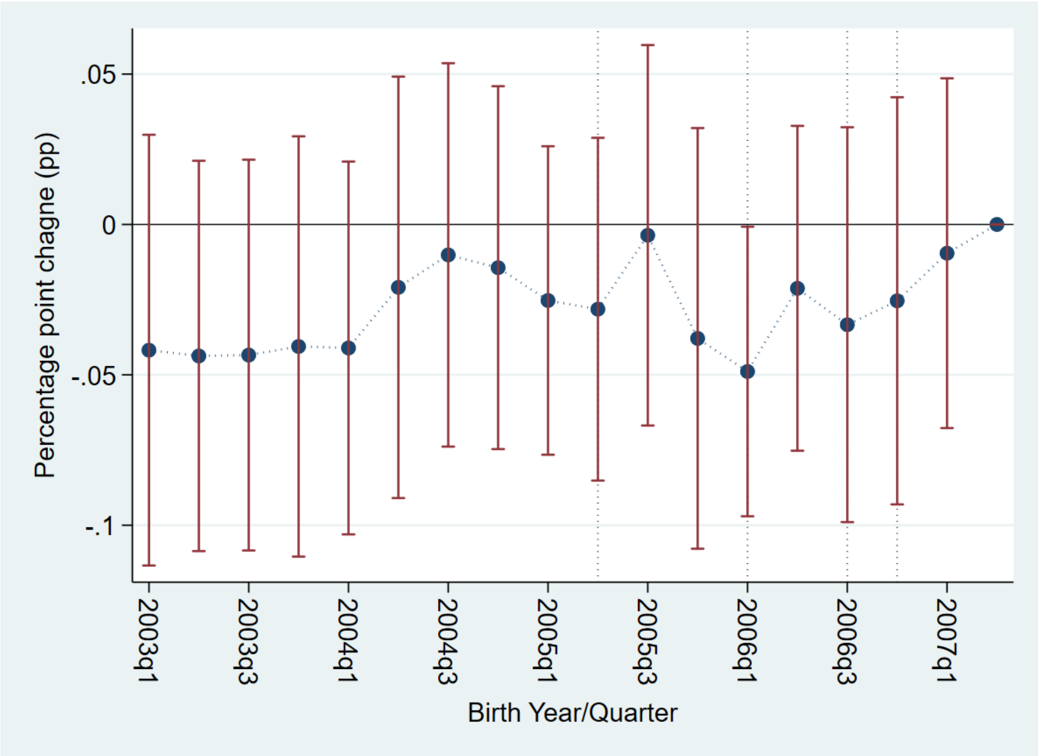
Notes: This figure shows the difference in mean outcomes of the treatment group for each birth year-quarter. Q2 2007 is set as a benchmark and equal to zero by construction. The vertical dotted lines show the year-quarter in which Family Start expanded into a new set of treatment TLAs.

Figure E7. Parallel trends test OT/CYF Intake event, year 1 (Māori)



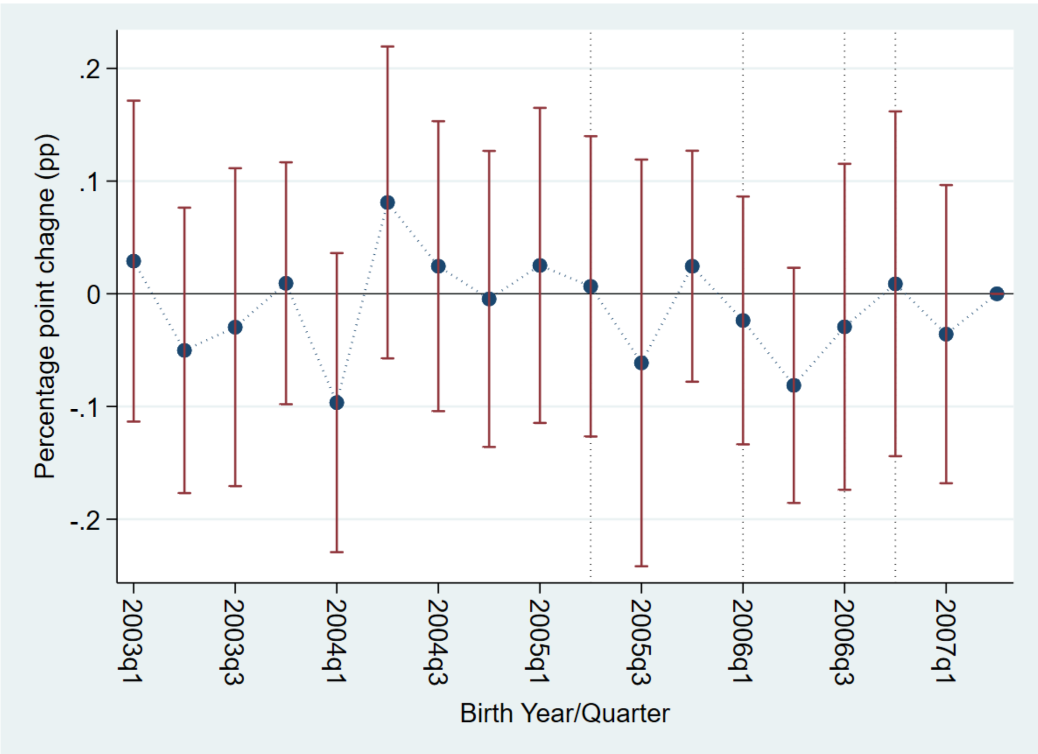
Notes: This figure shows the difference in mean outcomes of the treatment group for each birth year-quarter. Q2 2007 is set as a benchmark and equal to zero by construction. The vertical dotted lines show the year-quarter in which Family Start expanded into a new set of treatment TLAs.

Figure E8. Parallel trends test Family Violence event, year 1 (Māori)



Notes: This figure shows the difference in mean outcomes of the treatment group for each birth year-quarter. Q2 2007 is set as a benchmark and equal to zero by construction. The vertical dotted lines show the year-quarter in which Family Start expanded into a new set of treatment TLAs.

Figure E9. Parallel trends test Family Violence event, year 6 (Pasifika)



Notes: This figure shows the difference in mean outcomes of the treatment group for each birth year-quarter. Q2 2007 is set as a benchmark and equal to zero by construction. The vertical dotted lines show the year-quarter in which Family Start expanded into a new set of treatment TLAs.

Table E5. Joint significance test (*F*-test)

	Mean outcome	Test score
All children		
Maternal mental health and addiction service use – year 1	15.9%	**3.0182*** **
Maternal mental health and addiction service use – year 2	16.2%	**2.0899** **
Family Violence notifications – year 1	17.3%	1.18
Family Violence notifications – year 6	16.0%	**2.0409** **
Māori		
OT/CYF notifications – year 1	25.2%	**2.9593*** **
Family Violence notifications – year 1	19.4%	1.6082*
Pasifika		
Family Violence notifications – year 6	13.4%	**1.9856** **

Notes: Shares are based on randomly rounded counts in accordance with Stats NZ's random rounding 3 regulations. For a full list of variables, see Appendix C. Test scores refer to the results from joint significance tests, with estimates statistically different from zero at 99% (***), 95% (**), 90 (*) confidence.

DiD estimates using the 'benefit' inclusion criterion

This subsection presents the regression results from our main DiD estimation (Table 6 -Table 14), using the 'benefit' inclusion criterion. Note that these results are based on the mother-benefit inclusion criterion (unlike the child-benefit criterion used in the replication section above), as the IDI refresh used was older, and the error in the child benefit table had not been corrected.

Table E6. DiD outcomes, any ethnicity, benefit inclusion criterion (2003-2015)

	Outcome period		
	First year	Second year	Sixth year
Difference in Difference estimates			
Post neonatal mortality	-0.00083	-0.00085	0.00002
	[-0.00247, 0.00082]	[-0.00249, 0.00078]	[-0.00061, 0.00065]
Post neonatal injury death	-0.00050*	0.00011	0.00001
	[-0.00102, 0.00002]	[-0.00019, 0.00041]	[-0.00043, 0.00045]
Post neonatal SUDI	-0.00028		
	[-0.00151, 0.00095]		
OT/CYF Report of Concern or other initial intake event	**0.01126** **	0.00110	**-0.01083** **
	[0.00064, 0.02188]	[-0.01336, 0.01557]	[-0.02065, -0.00101]
OT/CYF assessment	0.00974*	0.00248	-0.00388
	[-0.00032, 0.01980]	[-0.00556, 0.01052]	[-0.00968, 0.00193]
OT/CYF care placement	0.00240*	0.00101	-0.00129
	[-0.00020, 0.00501]	[-0.00218, 0.00420]	[-0.00382, 0.00124]
Child or sibling Police FV report to OT/CYF	**0.01383** **	0.00563	**-0.01588*** **
	[0.00039, 0.02727]	[-0.01703, 0.02830]	[-0.02605, -0.00570]
Hospitalised for maltreatment related injury	0.00002	0.00069	-0.00023
	[-0.00116, 0.00120]	[-0.00141, 0.00278]	[-0.00297, 0.00252]
Hospitalised for long bone fracture	0.00063*		
	[-0.00001, 0.00128]		
Enrolled with a PHO	-0.01275	-0.00469	-0.00103
	[-0.03088, 0.00538]	[-0.01263, 0.00325]	[-0.00382, 0.00177]
Mother received publicly funded mental health services	**-0.02138*** **	**-0.01094*** **	0.00299
	[-0.03143, -0.01134]	[-0.01899, -0.00289]	[-0.00689, 0.01288]

Note: Parameter estimates statistically different from zero at 99% (***), 95% (**), 90% (*) confidence.

Table E7. DiD outcomes, Māori ethnicity, benefit inclusion criterion (2003-2015)

	Outcome period		
	First year	Second year	Sixth year
Difference in Difference estimates			
Post neonatal mortality	-0.00221	-0.00219	0.00065
	[-0.00512, 0.00069]	[-0.00503, 0.00065]	[-0.00048, 0.00177]
Post neonatal injury death	-0.00081	0.00020	0.00033
	[-0.00184, 0.00022]	[-0.00034, 0.00075]	[-0.00054, 0.00120]
Post neonatal SUDI	-0.00085		
	[-0.00343, 0.00172]		
OT/CYF Report of Concern or other initial intake event	0.02000***	0.00677	-0.01529*
	[0.00607, 0.03392]	[-0.00918, 0.02271]	[-0.03056, -0.00002]
OT/CYF assessment	0.01821**	0.00434	-0.00648
	[0.00364, 0.03279]	[-0.00528, 0.01395]	[-0.01470, 0.00173]
OT/CYF care placement	0.00268	0.00126	-0.00310
	[-0.00112, 0.00647]	[-0.00345, 0.00596]	[-0.00776, 0.00156]
Child or sibling Police FV report to OT/CYF	0.01882**	0.01532	-0.01264
	[0.00226, 0.03538]	[-0.01227, 0.04290]	[-0.03064, 0.00536]
Hospitalised for maltreatment related injury	-0.00032	0.00185	-0.00084
	[-0.00177, 0.00112]	[-0.00126, 0.00497]	[-0.00547, 0.00378]
Hospitalised for long bone fracture	0.00042		
	[-0.00060, 0.00143]		
Enrolled with a PHO	-0.00856	-0.00331	-0.00153
	[-0.02650, 0.00938]	[-0.01273, 0.00612]	[-0.00434, 0.00128]
Mother received publicly funded mental health services	-0.00808	-0.00550	-0.00449
	[-0.01792, 0.00175]	[-0.01371, 0.00270]	[-0.01655, 0.00758]

Note: Parameter estimates statistically different from zero at 99% (***), 95% (**), 90% (*) confidence.

Table E8. DiD outcomes, Pasifika ethnicity, benefit inclusion criterion (2003-2015)

	Outcome period		
	First year	Second year	Sixth year
Difference in Difference estimates			
Post neonatal mortality	0.00207	0.00198	-0.00009
	[-0.00097, 0.00511]	[-0.00111, 0.00507]	[-0.00147, 0.00129]
Post neonatal injury death	0.00042	0.00044	0.00040
	[-0.00129, 0.00214]	[-0.00069, 0.00157]	[-0.00055, 0.00134]
Post neonatal SUDI	0.00181		
	[-0.00073, 0.00436]		
OT/CYF Report of Concern or other initial intake event	-0.00221	0.00317	-0.00319
	[-0.01915, 0.01473]	[-0.01915, 0.02549]	[-0.02008, 0.01370]
OT/CYF assessment	0.00234	0.00168	-0.00626
	[-0.01242, 0.01711]	[-0.01101, 0.01533]	[-0.01771, 0.00520]
OT/CYF care placement	-0.00115	-0.00071	-0.00318*
	[-0.00554, 0.00323]	[-0.00584, 0.00441]	[-0.00682, 0.00047]
Child or sibling Police FV report to OT/CYF	0.00046	-0.00273	**-0.02255*** **
	[-0.01618, 0.01711]	[-0.02503, 0.01957]	[-0.03664, -0.00846]
Hospitalised for maltreatment related injury	**-0.00290** **	0.00299*	-0.00064
	[-0.00520, -0.00060]	[-0.00018, 0.00616]	[-0.00496, 0.00369]
Hospitalised for long bone fracture	-0.00042		
	[-0.00251, 0.00168]		
Enrolled with a PHO	-0.00706	0.00317	-0.00055
	[-0.02889, 0.01478]	[-0.00997, 0.01631]	[-0.00489, 0.00380]
Mother received publicly funded mental health services	-0.01275*	-0.00706	-0.00404
	[-0.02499, -0.00051]	[-0.01772, 0.00361]	[-0.01923, 0.01114]

Note: Parameter estimates statistically different from zero at 99% (***), 95% (**), 90% (*) confidence.