



**ORANGA  
TAMARIKI**  
Ministry for Children

**EVIDENCE CENTRE**  
TE POKAPŪ TAUNAKITANGA

New Zealand Government

# Reducing Youth Reoffending in South Auckland Social Bond pilot

## Quantitative Impact Evaluation

**Email:** [research@ot.govt.nz](mailto:research@ot.govt.nz)

**Authors:** Eyal Apatov and Philip Spier (Oranga Tamariki Evidence Centre)

**Published:** June 2024

**ISBN:** 978-1-7386280-7-0

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

### **Citation guidance:**

Apatov, E., & Spier, P. (2024). *Reducing Youth Reoffending in South Auckland Social Bond pilot: Quantitative Impact Evaluation*. Wellington, New Zealand: Oranga Tamariki—Ministry for Children.

### **Copyright:**

This document *Reducing Youth Reoffending in South Auckland Social Bond pilot: Quantitative Impact Evaluation* is licensed under the Creative Commons Attribution 4.0 International License <http://creativecommons.org/licenses/by/4.0/>. Please attribute © New Zealand Government, Oranga Tamariki—Ministry for Children 2024.

### **Acknowledgements**

We would like to thank Dean Hyslop and Trinh Le (Motu Economic and Public Policy Research) for advice at all stages of this project and review of drafts and code, and Sarah Crichton (New Zealand Treasury) for review of analysis and for the many helpful discussions and suggestions. We would also like to thank our Oranga Tamariki colleagues Colin Hewat (Evidence Centre), Lela Leong (Māori Partnerships and Communities), and Josh Chester-Master (Service Delivery) for their review and feedback, and to Freya Bacon-Bootham (Evidence Centre) for proofreading and editing.

### **Disclaimer – Oranga Tamariki**

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.

### **Disclaimer – Statistics New Zealand**

These results are not official statistics. They have been created for research purposes from the Integrated Data Infrastructure (IDI) which is carefully managed by Stats NZ. For more information about the IDI, please visit:

<https://www.stats.govt.nz/integrated-data/>



# Contents

<b>Non-technical summary</b> .....	<b>1</b>
<b>Introduction</b> .....	<b>5</b>
Background .....	5
Related literature .....	8
<b>Study design</b> .....	<b>12</b>
Data.....	12
Study populations.....	13
Matching.....	15
Limitations .....	22
<b>Findings</b> .....	<b>24</b>
Reoffending by Social Bond pilot participants .....	24
Estimated impacts of the pilot.....	29
<b>Discussion</b> .....	<b>30</b>
<b>References</b> .....	<b>32</b>
<b>Appendix A</b> .....	<b>35</b>
<b>Appendix B</b> .....	<b>42</b>
<b>Appendix C</b> .....	<b>49</b>



## Tables and figures

Table 1: Summary of rangatahi characteristics, by referral/enrolment status, September 2017 to August 2023.....	7
Table 2: Reoffending frequency, 31 August 2022 .....	8
Table 3: Treatment group restrictions, sample sizes, and mean characteristics .....	14
Table 4: Number and share of final treatment group by (grouped) days between nearest prior proceeding and enrolment.....	14
Table 5: Distinct rangatahi and proceedings counts by restriction, comparator pool	15
Table 6: Matching variables .....	19
Table 7: Balance table, treatment group (SB participants) and matched control group .....	20
Table 8: Whether SB participants committed an offence of each type, and number of offences recorded by police, 12 months before and after enrolment.....	27
Table 9: Whether SB participants committed an offence within each seriousness band, and number of offences recorded by police, 12 months before and after enrolment .....	28
Table 10: Reoffending outcomes for SB participants in the 12 and 24 months after enrolment .....	28
Table 11: Social bond pilot impact estimates by outcome measure and reoffending window .....	29
Table 12: Social bond pilot impact estimates for rangatahi Māori by outcome measure and reoffending window .....	29
Table 13: Social bond pilot impact estimates for non-Māori by outcome measure and reoffending period .....	30
Table A1: Participant characteristics .....	35
Table A2: Frequency of offences, occurrences, and proceedings.....	36
Table A3: Relationship between matching variables and 12-month outcomes, all rangatahi with AA diversion during pilot period .....	39
Table B1: Reoffending outcomes for SB participants in the 12 months after enrolment, by gender .....	42
Table B2: Reoffending outcomes for SB participants in the 12 months after enrolment, by ethnicity (prioritised) .....	42
Table B3: Reoffending outcomes for SB participants in the 12 months after enrolment, by age at enrolment.....	42
Table B4: Mean and differences in outcomes by specification.....	43
Table B5: Balance table, various specifications .....	45
Table B6: Balance table by number of nearest neighbours.....	47
Table C1: Social bond pilot impact estimates by outcome measure and reoffending period, by (minimum) number of neighbours.....	49
Table C2: Social bond pilot impact estimates by outcome measure and reoffending period, using reference-proceeding as reference date for participants .....	50
Table C3: Social bond pilot impact estimates by outcome measure and reoffending period, excluding rangatahi who were prosecuted at reference proceeding.....	50
Table C4: Social bond pilot impact estimates by outcome measure and reoffending period, all rangatahi with AA outcome at reference proceeding .....	50
Table C5: Social bond pilot impact estimates by outcome measure, reoffending period. and maximum lifetime offence severity .....	51

Table C6: Social bond pilot impact estimates by outcome measure and reoffending period, all rangatahi processed in Counties Manukau Police District at reference proceeding .....	51
Table C7: Social bond pilot impact estimates by outcome measure and reoffending period for rangatahi who offended within 12 and 24 months .....	52
Table C8: Social bond pilot impact estimates by outcome measure and reoffending period, alternative outcome measures .....	52
Table C9: Propensity Score Match estimates, all participants.....	55
Table C10: Inverse Probability Weighted Regression Adjustment estimates, all participants.....	55
Figure 1: Quarterly referrals and enrolments.....	8
Figure 2: Cumulative percentage of SB participants who had reoffended over the 12 months following enrolment.....	24
Figure 3: Average frequency of offences by SB participants in each six-month period before and after enrolment .....	25
Figure 4: Average total seriousness of offences by SB participants in each six-month period before and after enrolment .....	26
Figure A1: Annual distinct offences, occurrences, proceedings, and individuals, 2009-2023 .....	37
Figure C1: Total offending distribution by treatment status and reoffending window	53
Figure C2: Maximum offence severity distribution by treatment status and reoffending window .....	54
Figure C3: Total offence severity distribution by treatment status and reoffending window .....	54

# Non-technical summary

## Introduction

Social Impact Bonds (or Social Bonds) are a type of pay-for-success contracting approach that have been used internationally to fund a variety of social interventions, including youth recidivism prevention. Under this approach, the government contracts private investors to fund a social service (via a service provider), paying the investors back, plus a return, if pre-determined outcomes set in the contract are met.

This report focusses on the only bond successfully launched to date in New Zealand, the *Reducing Youth Reoffending in South Auckland Social Bonds Pilot* (SB pilot). The goal of this bond was to reduce the frequency and severity of youth reoffending in South Auckland. Capped at NZD\$24m, the pilot was operated by Genesis Youth Trust (GYT), who were tasked to deliver an intervention that will improve the offending outcomes of up to 1,000 (mostly) South Auckland children and young people (rangatahi henceforth) who offended and had a medium to high risk of reoffending.

Oranga Tamariki was responsible for managing the bond's Outcome Agreement and for evaluating the pilot.

This evaluation utilises data up to June 2023 from Police and the Youth Justice National Minimum Dataset (YJNMDs) to quantify the impacts of the pilot on participants' reoffending outcomes. The findings complement those in two previous evaluations Oranga Tamariki commissioned on the SB pilot.<sup>1</sup>

## What was the SB pilot?

Rangatahi were referred to the pilot by New Zealand Police, and if enrolled, they took part in a 20-week intensive programme, followed by ongoing mentoring and monitoring for up to 18 additional months. In total, 804 rangatahi were referred to the pilot, and 607 (75%) were successfully enrolled. Most participants in the pilot identified as Māori, males, approached the age of 15 at referral, and recorded a medium risk of reoffending (an average score of 55 in the Police YORST tool, which assesses the risk of youth reoffending).

In terms of performance, outcomes were assessed every six months, with payments based on whether changes in outcomes for participants outperformed contracted targets. Specifically, these were linked to achieving or outperforming pre-determined YLS/CMI assessment scores,<sup>2</sup> as well as reductions in reoffending frequency and severity targets.

---

<sup>1</sup> For more information, see Malatest International and Oranga Tamariki Evidence Centre (2021) and Allen & Clarke (2023).

<sup>2</sup> The Youth Level of Service/Case Management Inventory (YLS/CMI) tool measures the needs and risk of youth (e.g., family and living circumstances, education and employment, peer relations, substance abuse, attitudes, and beliefs) to assess their risk of reoffending. For more information, see: [Youth Level of Service Inventory \(YLS 2.0\) Ratings & User Guide \(pa.gov\)](#).

## Reoffending outcomes and matching variables

To assess pilot effectiveness, participants reoffending frequency and severity outcomes were assessed across 12- and 24-month windows from enrolment. The impact of the pilot on **reoffending frequency** was assessed by whether rangatahi reoffended, and whether the total number of offences post-enrolment date was lower than the total number of offences in the equivalent period of months leading up to enrolment. The impact of the pilot on **reoffending severity** was assessed by whether the most severe offence as well as the combined severity across all offences, were lower in the period following enrolment than in the equivalent period leading up to it.

The evaluation focused on the outcomes of the 471 and 382 participants who were enrolled for at least 12- and-24-months respectively at the data cut-off date (June 2023). The impact of the pilot on participant outcomes was estimated relative to those experienced by Auckland-based rangatahi with statistically similar or identical observable characteristics, but who were not enrolled to the pilot. Similarity was measured in terms of a selection of demographic and prior (largely) Justice sector related interactions (using Nearest Neighbour Matching).

## Participant characteristics

Participants tended to record a small number of offences and incidents before enrolment, with 40% recording a single offence, and more than 80% up to three offences. Generally, the frequency and seriousness of offending had escalated within the six months leading to enrolment, with most not offending much (or at all) in earlier periods. Illegal use of a motor vehicle, property damage, injury causing acts, and public order/weapon related offences were some of the most common types of offences.

Over 80% of participants recorded Alternative Action (AA, a police youth diversionary response to offending) as the proceeding closest to enrolment. Only a small fraction recorded more escalated Police outcomes (e.g., prosecution). Next, 70% recorded their first-ever offence between the ages of 10 and 14, about one-third were females, over 70% identified as Māori, and 28% identified as Pacific peoples (using total response ethnicity). Around 30% of participants recorded prior Care and Protection interaction with Oranga Tamariki (e.g., investigation, care placement).

In terms of similarity between participants and the matched group of non-participants, most matching variables used showed similar means and distributions. However, SB participants recorded relatively more severe offences prior to the reference date, speculatively downward biasing (i.e., *understating*) the evaluation's estimated impact. In contrast, the matched group of non-participants recorded a relatively larger number of prior police incidents/tasks, and (1-3pp) more recorded a prosecution at the reference proceeding. Speculatively, this may have had the opposite effect (i.e., *overstating*).

## Participant outcomes

In the 12 months following enrolment, over half (54%) of the participants did not reoffend, and around 70% reduced the frequency and/or seriousness of their

offending (compared to the 12 months before). In the 24 months following enrolment, 46% of participants did not offend, and about two-thirds reduced the frequency and/or seriousness of their offending (compared to the 24 months before).

Typically, the share and number of participants who committed an offence fell in the 12 months following enrolment (compared to the 12 months before), though the total number of offences increased by 6%. A similar pattern was observed when grouping offending into seriousness bands. That is, the number and share of participants who offended decreased within each seriousness band, while the total number of offences in the three highest bands increased. Combined, these findings indicate that a small sub-group of participants recorded high numbers of offences (with greater total severity) following enrolment, while most other participants recorded a decline or did not reoffend. Much of this was driven by changes in the number of illegal use of motor vehicle, burglary, and robbery-related offences.

### **SB pilot's estimated impacts**

Our main findings indicated that the pilot successfully reduced the frequency of participant reoffending. In the 24-months following enrolment, the likelihood of participants not reoffending increased by 28% compared to the matched control group. Put differently, the estimates suggest that if participants would have been referred to the same business-as-usual type interventions as the group of matched non-participants (rather than to the SB pilot), 61 additional participants would have reoffended within 24 months. In addition, participants were 14% and 24% more likely to reduce the frequency of their offending in the 12- and 24-months post enrolment, respectively. While indicative of positive change, the impacts relating to reoffending severity were not statistically significant.

On average, participants were estimated to commit nearly two (1.976) fewer offences in the 24-months following enrolment, which in turn, aggregates to a total of nearly 1,200 fewer offences. Additional exploration of the data indicates that fewer offences were driven both by a greater share of participants not reoffending, as well as by fewer offences committed by those who did.

Our main findings were largely repeated when estimating the impacts for different sub-groups (e.g., ethnicity, alternative reference date, reference proceedings' outcome, location, prior offence severity), when alternating the algorithm of the model, and when applying alternative matching techniques. This was especially true in terms of consistently estimating large, positive, and significant reductions in participant reoffending frequency in the 24 months following enrolment.

### **Future analysis**

Given that at the time of the evaluation, not all participants had completed the two-year intervention (or had two years of offending data available post-enrolment), a reasonable next step would be to re-examine the pilot impacts in the future. A supplementary benefit of this would be the ability to explore longer-term impacts (i.e., beyond two years).

In addition, a simplistic approximation that monetises the aggregated reduction in offending (1,200 fewer offences) by the average cost of an offence (\$15,487),



indicates that the pilot may have led to savings of between 18 and 19 million NZD in the two years following enrolment. Given such large potential savings over a relatively short period of time, we recommend formal assessment of the pilot's return on investment via a Cost Benefit Analysis (CBA). While this will require additional work, including possible changes to the approach taken in this evaluation, a formal CBA will improve our understanding of whether this pilot specifically, and the Social Impact Bonds approach more broadly, could complement (or be an alternative to) the standard business as usual type approaches that are in use for contracting social services.

# Introduction

Social Impact Bonds (or Social Bonds) are a type of Pay-for-Success contracting approach that have been used to fund a variety of social interventions, including youth recidivism prevention (Ramsen, 2016). By May 2024, 292 Social Bonds have been launched across 38 countries, with a combined worth of over US\$764m.<sup>3</sup>

In this approach, the government contracts private investors to fund a social service (via a service provider) and pays the investors back, plus a return, if predetermined outcomes in the contract are met (Ministry of Health, 2015). The rationale for adopting this contractual approach includes shifting the financial risk of service delivery from the public sector to private investors, providing a clear financial incentive to achieve measurable outcomes, allowing providers more freedom to innovate by having a more hands off approach, and typically, increasing financial certainty for providers to budget and plan via multi-year funding agreement (Ministry of Health, 2015; Ramsen, 2016; Minister of Finance & Minister Responsible for Social Investment, 2017).

In September 2017, Oranga Tamariki entered a six-year agreement with Genesis Youth Trust (GYT)<sup>4</sup> to deliver an intervention aimed at reducing reoffending severity and frequency of South Auckland rangatahi who came to the attention of police for offending. The *Reducing Youth Reoffending in South Auckland Social Bond* pilot (SB pilot) included five years of enrolment into the pilot, as well as an additional year for delivering the intervention to remaining rangatahi participating in the pilot after the enrolment cut-off date (31 August 2022).

This report evaluates the impact of the SB pilot by estimating different participant reoffending outcomes relative to those experienced by a matched group of rangatahi who did not partake in the pilot.

## Background

The SB pilot was funded by the New Zealand Superannuation Fund, Mint Asset Management (a private fund manager) and the Wilberforce Foundation (a private philanthropic investor) who provided an initial investment of NZD\$6m. Payments were linked to outcomes, capped at NZD\$24m over the pilot's period for up to 1,000 participants.

---

<sup>3</sup> For more information, see [Impact Bond Dataset \(ox.ac.uk\)](https://www.ox.ac.uk/impact-bond-dataset).

<sup>4</sup> GYT is a police-affiliated youth development organisation working with at-risk youth and their families across Auckland to reduce youth reoffending. It is a charitable trust and a hybrid organisation made up of six Police paid staff and 39 Trust paid personnel (Malatest International & Oranga Tamariki Evidence Centre, 2021).

The pilot targeted children and young people who offended, recorded a Youth Offending Risk Screening Tool (YORST) score<sup>5</sup> of between 40 and 100, and resided within the pilot's geographic boundary (mostly South Auckland suburbs).<sup>6</sup>

Rangatahi were referred to the pilot by Police with an Alternative Action (AA) plan.<sup>7,8</sup> Once referred, rangatahi and their whanau were required to sign a consent form (i.e., enrolment was voluntary), and an initial assessment was completed via the Youth Level Service/Case Management Inventory (YLS/CMI) tool.<sup>9</sup> During the enrolment process, in addition to not meeting criteria, GYT could decline participation if the rangatahi did not engage with staff, breached their AA plan, posed a risk to staff or other clients, or reoffended at a level where a more serious youth justice response was required. In addition, the contract required GYT to have at least 30% of their enrollees with a YORST score of 60 or above. Therefore, while it was theoretically possible for GYT decline referred rangatahi to maintain this share, there is no evidence that declining for this reason was required. Once enrolled, rangatahi took part in a 20-week intensive programme,<sup>10</sup> followed by ongoing mentoring and monitoring for a period of up to 18 additional months (i.e., as needed).

The first rangatahi was referred to the pilot in September 2017, with the first successful enrolment recorded in the following month. Table 1 presents the characteristics of all rangatahi who were referred and/or enrolled to the pilot. In total, 804 rangatahi were referred to the pilot, and of those, 607 (75%) successfully enrolled. The main reasons for rangatahi not being accepted to the pilot included lack of engagement (52%), reoffending that led to escalation to a family group conference (FGC) or prosecution during the enrolment process (18%), or not meeting the eligibility criteria (28%).

Of enrolled participants, 23% (140) left the pilot prematurely. Of those, nearly three-quarters could not be engaged/contacted with, one quarter due to reoffending, and one tenth due to rangatahi relocation outside the pilot's geographic boundary. When the pilot's enrolment period ended (31 August 2022), 21% of participants were in the

---

<sup>5</sup> The YORST score is the estimated risk of reoffending (e.g., a YORST score of 60 means the tool estimated the likelihood of the rangatahi reoffending was 60%). For more information about the use of YORST score by New Zealand Police, see Mossman (2016).

<sup>6</sup> For the purposes of the pilot, South Auckland was defined as the suburbs of Mangere, Otahuhu, Papatoetoe, Otara, Onehunga, Mt Wellington, Glen Innes, Panmure, Orakei, Manurewa, Clendon, Takanini, Papakura, and Pukekohe. In practice, it is not clear how the boundaries of these suburbs were defined by Police.

<sup>7</sup> When Police apprehend a child (aged 10 to 13) or a young person (aged 14 to 17) for an alleged offence, the case is referred to the Police Youth Aid Section to decide on how to deal with the case. One option when the offending has been admitted is to arrange an informal youth diversionary response known as Alternative Action (AA). This involves the development of a plan that will often include elements with the aim of making the young person accountable for his or her actions, making amends to the victim, and addressing criminogenic needs that the young person may have. Plans are developed by a Youth Aid officer in consultation with the young person and their parents or caregivers and will have a clearly defined timeframe for completion (Oranga Tamariki, 2018). For more information about AA, see McLaren (2010).

<sup>8</sup> According to the outcome agreement, a cross-agency multi-disciplinary youth offending team was required to determine whether the client was suitable for the pilot, and whānau and/or other significant members to the client must have agreed for them to attend.

<sup>9</sup> YLS/CMI is a risk/needs assessment and a case management tool. It identifies a youth offender's needs strengths, barriers, and incentives, to develop an effective case management plan: [multi-health-systems-usd \(mhs.com\)](https://mhs.com)

<sup>10</sup> Responses were structured around multidisciplinary teams that included social workers, counsellors, and youth workers/mentors. Once enrolled, the team prioritised addressing the immediate needs of the participant (e.g., food security), followed by a response tailored to address their specific immediate and longer-term needs. For more information, see Allen & Clarke (2023).

pilot's initial intensive phase (first 20 weeks), 20% were in the mentoring/monitoring phase (6 to 18 months), and 58% had completed the pilot.

Across all four groups in the table, most rangatahi were Māori, male, approaching the age of 15 at referral, and had a mean YORST score of about 55. Of those enrolled, one-third recorded a high YORST score (60-100). Across groups, females were slightly more likely to be declined (3 percentage points, henceforth pp) but also less likely to prematurely leave if enrolled (8pp less). In addition, older rangatahi, those with high YORST scores (4-5pp), and Pacific peoples, were all more likely to exit the pilot prematurely, with the first two also being more likely to be declined.

*Table 1: Summary of rangatahi characteristics, by referral/enrolment status, September 2017 to August 2023*

	Referred	Declined	Enrolled	Premature exit
Rangatahi numbers	804	173	607	140
Female	30%	33%	30%	22%
Average age at referral	14.5	14.9	14.1	14.7
Māori	70%	70%	70%	71%
Pacific peoples	18%	18%	19%	24%
Mean YORST at referral	55.1	57.3	54.6	55.2
Share with YORST 60+	35%	42%	33%	36%

**Source:** Authors' calculations from data provided by Synergia (2023). **Notes:** As of September 2023. Referred rangatahi includes 24 whose enrolment to the pilot was still pending at the time the data was extracted.

Figure 1 shows quarterly referrals and enrolments to the pilot during its lifetime. The figure shows greater quarter-to-quarter variation in enrolments than referrals. There were fewer referrals and enrolments from 2020, due at least in part to the national lockdowns following the COVID-19 outbreak in 2020, and the Auckland-based lockdowns in the following year. For example, average quarterly referrals and enrolments up to 2020 were 44 and 35 respectively but fell to 37 and 27 afterwards. In addition, the average time to enrol rangatahi into the pilot (i.e., measured as days between the referral and enrolment dates) increased by 75% from 33 days in 2017–2019, to an average of 58 days from 2020.

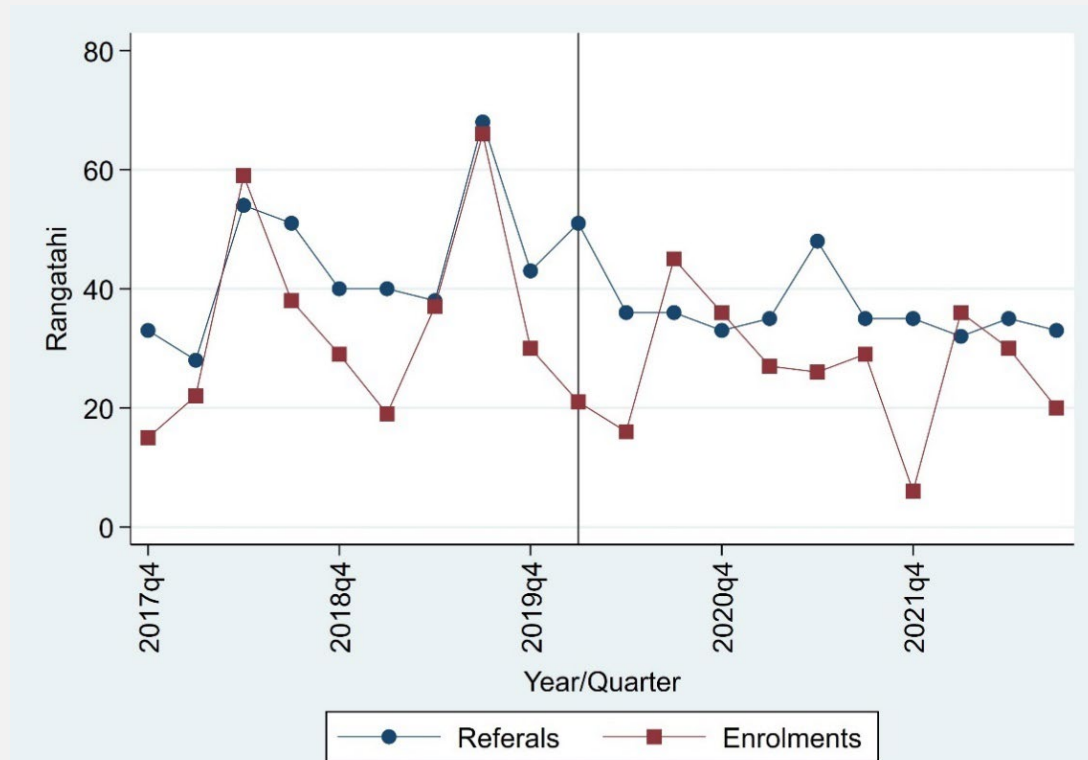
In terms of performance, outcomes were assessed every six months, with payments based on whether changes in outcomes for participants outperformed various pre-agreed targets (termed 'counterfactuals' in the contract).<sup>11</sup> The outcomes in the contract were measured by (YORST cohort level) changes in YLS/CMI scores, reoffending frequency, and reoffending severity. Note that GYT staff internally completed the YLS/CMI assessments (which were then subject to external auditing), while offending history data were provided by Police and analysed by Synergia. For example, Table 2 presents participant reoffending frequency as of August 2022.<sup>12</sup> The table is restricted to pilot participants who enrolled at least 24 months prior to this date and shows both the actual reoffending rate per participant for each YORST category, and the counterfactual target in brackets. The table shows that in nearly all

<sup>11</sup> These targets varied based on the outcome measured, YORST cohort (medium, high), and period horizon (first 6 months since enrolment to 18–24 months post enrolment). The counterfactual measured the same reoffending outcomes as recorded for rangatahi who participated in an GYT intervention that operated between 2009 and 2017 (12-week intervention that the SB pilot was an extension of).

<sup>12</sup> Note that in practice, outcomes were measured by enrolment cohort, and over 6-month windows. That is, the table presents these outcomes for illustrative purposes.

periods, the groups achieved a rate that was below (better than) the counterfactual. Similar results were found when measuring changes in the two other outcomes (Allen & Clarke, 2023). Note that outcomes for the purpose of assessing performance were for all enrollees, including participants who prematurely left the pilot (i.e., including those who exited due to reoffending).

Figure 1: Quarterly referrals and enrolments



Source Synergia (2023). Notes: Values from 2017q3 were combined with those from 2017q4 due to small numbers. Vertical black line indicates the first national lockdown following the COVID-19 outbreak.

Table 2: Reoffending frequency, 31 August 2022

YORST category	Participants	0-6m	>6-12m	>12-18m	>18-24m
Medium	274	0.88 (1.08)	0.43 (0.96)	0.54 (0.82)	0.4 (0.94)
High	115	1.93 (1.29)	0.83 (1.59)	0.82 (1.32)	0.97 (1.11)
<b>Total</b>	<b>389</b>	<b>1.19</b>	<b>0.54</b>	<b>0.62</b>	<b>0.57</b>

Source: Allen & Clarke (2023). Notes: Analysis limited to enrolled rangatahi who had complete monitoring data for at least 24 months post-enrolment. This included rangatahi who exited the pilot early, but for whom reoffending data were still captured who completed by YORST grouping (medium, high) and time since enrolment as of August 2022.

## Related literature

Internationally,<sup>13</sup> Social Bonds aimed at reducing youth recidivism have been implemented in the United States, Canada, Portugal, the United Kingdom, and Australia, including the first bond ever issued. This first bond was launched in 2010 in Peterborough (England) for an intervention designed to prevent recidivism for male youth with a short prison sentence by providing various forms of support (e.g.,

<sup>13</sup> This section is largely based on the findings from Le et al. (2022).

mentoring, education, job training). Using Propensity Score Matching (PSM), two separate evaluations (Jolliffe & Hedderman, 2014; Anders & Dorsett, 2017) found that on average, the reoffending rate of participants reduced by more than 9% (relative to the matched control group), well exceeding the 7.5% contracted target.<sup>14</sup>

In contrast, a bond in New York City aimed at reducing juvenile recidivism failed to meet the 10% reduction target set in the contract and was discontinued from the end of August 2015 (MDRC, 2015). In more detail, this intervention targeted male youth (aged 16-18) who entered Rikers Island jail for more than four days. To evaluate its impact, Parsons et al. (2016) used a PSM approach and matched eligible youth with youth who had similar characteristics (e.g., charge, criminal history, gender, age) but were now too old to be eligible (passed through jail between 2006 and 2010).<sup>15</sup> In the first year, the recidivism rate for the treatment group was not statistically different compared to that from the control.

More generally, despite a growing interest in this contracting approach globally, only a relatively small number of interventions undergo an impact evaluation that specifically focuses on isolating the effects of the interventions (Ten et al., 2021). For example, Fox and Morris (2021) reported that evaluations of many United Kingdom social bonds are largely qualitative, and/or had no rigorous attempt to establish a counterfactual, while Gustafsson-Wright et al. (2015) found that only four evaluations out of 38 reviewed used experimental methods (and an additional four used quasi-experimental methods).

Therefore, it is not yet clear whether many of the bonds that were issued globally effectively meet their goals (i.e., as opposed to meeting their contracted targets), and are more effective at addressing social issues than more *traditional* financing approaches for social initiatives (Wang, 2022).

### Previous evaluations of the SB pilot

In New Zealand, the SB pilot was evaluated three times. Malatest International & Oranga Tamariki Evidence Centre (2021) conducted a process evaluation once the pilot had been operating for at least two years. The evaluation highlighted several challenges that prevented the service from being delivered more effectively. These included complex and time-consuming procurement and contractual processes, overreliance on staff relations with specific Police staff for receiving referrals (due to Police not having a clear referral pathway process),<sup>16</sup> and the inability to review the contractual outcome measures during the pilot's duration.<sup>17</sup> Despite these challenges, analysis of the pilot's operational data found reductions in reoffending outcomes that surpassed the pilot counterfactuals (for all participants, and for rangatahi Māori and non-Māori). GYT staff mentioned several benefits from

---

<sup>14</sup> In these studies, bond participants were matched to up to 10 offenders based on similar characteristics and number of reconvictions.

<sup>15</sup> To control for factors that affect recidivism but are unrelated to the programme, such as city-wide changes in crime rates or changes in policing practices, the trend in recidivism for 19-year-olds is used to adjust the results. Nineteen-year-olds were chosen for this purpose because they were not eligible to participate in the programme yet historical data show that trends in recidivism for 16- to 18-year-olds and 19-year-olds were similar.

<sup>16</sup> This may (at least partially) explain why enrolments were lower than the maximum target set by the contract.

<sup>17</sup> Due to the small size of the cohort and reoffending outcomes measured at the cohort level, outliers could disproportionately affect the overall outcomes.

operating under the Social Bond contractual approach, including greater accountability to investors (which in turn, incentivised them to perform at their best), multi-year financing that introduced certainty regarding finance and supported longer-term investment and planning decisions (e.g., workforce development), and the use of a structured and evidence-based practice model (the YLS/CMI tool).

Next, Allen & Clarke (2023) focused on the experiences of participants, their whānau, as well as GYT staff and other key stakeholders as part of a qualitative outcomes evaluation. Overall, pilot participants reported that they felt supported by GYT, and that this support helped them to achieve their aspirations and (long-lasting) positive outcomes (e.g., improved whānau relationships and communication, educational goals, not reoffending). In addition, the pilot was reported to also be assisting participant's whānau to think about their own aspirations and set and achieve positive goals.<sup>18</sup> Using operational data over a longer period, Allen & Clarke (2023) also identified reductions in all key outcome measures as participants progressed through the programme, with most outcomes outperforming the contracted targets (counterfactuals).

Next, Synergia (2023) utilised YORST data added to the Integrated Data Infrastructure (IDI) to examine effects of the pilot on participant outcomes.<sup>19</sup> For this, participant outcomes were compared with those experienced by non-participants that were matched by age, gender, ethnicity, and YORST assessment related factors (total score, assessment year, Auckland Police Districts processing the assessment). While the data included a flag to indicate SB pilot participation, a limitation of the analysis was that it did not include pilot enrolment dates. To attempt to address this gap, Synergia instead used the earliest date in which rangatahi recorded a medium or high YORST score (40 or more), as the reference date for measuring outcomes from. Overall, the evaluation found that relative to the matched comparison group, participants recorded fewer offences and criminal charges, were less likely to consume prescriptions and mental health services, and were more likely to enrol in tertiary education, and recorded higher incomes. In a subsequent evaluation, Bakker (2023) monetised these impacts, estimating a benefit-cost-ratio of 2:1 (i.e., two dollars benefit for each dollar invested) within six years post-enrolment, and 12:1 over the participants' lifetime.

Despite the work completed to date, uncertainties regarding the effectiveness of the pilot remain. For example, while both Malatest International & Oranga Tamariki Evidence Centre (2021) and Allen & Clarke (2023) identified reductions in participant reoffending, both evaluations stated that these reductions cannot be attributed to the pilot due to the lack of a robust counterfactual.<sup>20</sup> In addition, while Synergia (2023) used matching to isolate the impact of the SB pilot, the methodology used consists of some features that compromise the reliability of its findings. Specifically, it is not

---

<sup>18</sup> In terms of the bond structure, staff mentioned similar benefits of operating under this approach to those recorded by Malatest International & Oranga Tamariki Evidence Centre (2021).

<sup>19</sup> The IDI is a data warehouse that includes linked individual level data from a range of government agencies and sources (e.g., Inland Revenue, Ministry of Education, Ministry of Health). For more information on the IDI, see: <https://www.stats.govt.nz/integrated-data/integrated-data-infrastructure>

<sup>20</sup> In more detail, since contracted targets were based on past reoffending trend, and constant (e.g., 1.08 offences per participant with medium YORST during the first six months post-enrolment), they did not adjust to reflect broader youth justice trends. For example, Ministry of Justice (2024) reported that between 2010/11 and 2021/22, the offending rate of young people fell by 63%, and that by 2019/20, the 24-month reoffending rate for young people who were diverted to Alternative Action (or received a Warning) fell by nearly 14%.

clear whether the use of first medium to high YORST was a valid proxy for enrolment date, for example when participants recorded multiple YORST assessments with medium to high scores well before enrolment. Furthermore, data provided by Synergia shows that the enrolment date of some participants was months following their YORST assessment, and it is also unclear whether the assessment date is a reliable proxy for non-participants' business-as-usual intervention date (such as when an AA plan was put in place).

In addition, participants were largely enrolled with an AA plan, and Police are only systematically required to do a YORST assessment for rangatahi who are to be escalated to an FGC or be prosecuted. As a result, participants were largely matched with non-participants who recorded more severe types of Police intervention. Therefore, if (largely AA) participants and non-participants (largely FGCs or prosecution) had any unobserved systematic differences that correlated with outcomes that were not fully controlled by their YORST score, then the estimated reductions will be biased, even after matching is applied. For example, Oranga Tamariki (2018) found that the characteristics of rangatahi who received AA were too different to rangatahi who received an intention-to-charge FGC for making meaningful comparisons of outcomes (lack of common support under a Propensity Score Matching approach). Furthermore, when focusing on a smaller subset of rangatahi with similar observable characteristics, the study found a six-percentage point (pp) greater 12-month reoffending rate amongst a cohort of rangatahi who were referred to an FGC compared to a cohort diverted to AA (45% compared to 39%). That is, the study found differences in reoffending patterns even after matching a far larger and comprehensive set of variables (but not the total YORST score) than used in Synergia (2023).<sup>21</sup>

With regards to the YORST tool, while it was validated to be suitable to predict reoffending risk in New Zealand (Mossman, 2016),<sup>22</sup> it is not clear whether it is sufficiently accurate for the purpose of matching at the individual level. For example, Mossman (2016) reported that total score as a predictor of re-apprehension was less accurate for Māori, young offenders, as well as being less accurate in distinguishing between medium (score of 30-69) and high (70-100) risk offenders (than between low (0-29) and medium risk offenders). In addition, while all 14 YORST items that comprise the total score were correlated with re-apprehension, they differed in terms of their importance, with some items being redundant (not providing any added value). Put together, it is not clear whether matching on individual's *total* YORST score in Synergia (2023) was sufficient to fully control for the risk of reoffending between the treatment and control groups, especially since a large proportion of the sample included rangatahi Māori, all had medium and high YORST scores, and there was a potentially different distribution of underlying items (since most participants had an AA while most controls were referred to an FGC or were prosecuted).

---

<sup>21</sup> The sample include all AA and FGC recipients between July 2010 and June 2014. Matching was done using Propensity Score Matching, with gender, ethnicity, age at intervention, age at first offence, number of offences (prior 12 months, any period), total seriousness of offences (past 12 months, all past offences), and total offences by offence category (injury causing, theft-related, property damage, burglary, public order) as matching variables.

<sup>22</sup> Aggregately, total YORST score was found to correctly predict re-apprehension for 68% of cases examined.



# Study design

## Data

### *SB pilot participants*

New Zealand Police (Police) data was sourced to identify SB pilot participants. The data received from Police included the unique personal identifier from the National Intelligence Application (NIA) for rangatahi who were identified as participants by having a NIA 'alert' of "Genesis Social Bond Client". The alert creation date was also provided, and this was usually the pilot's enrolment date. Police kept a systematic record of participants since they were required to provide offending data to Synergia to support the periodic assessment of the pilot's performance (and to determine payments) related to the frequency and severity of reoffending. Overall, this source includes unique identifiers and start dates for 603 of the 607 rangatahi who participated in the SB pilot. Data regarding rangatahi who were referred to the pilot but did not enrol was not systematically collected by Police.

### *Offending histories*

Records of offending of rangatahi up to 30 June 2023 were sourced from the Police proceedings table within the Youth Justice National Minimum Dataset (YJNMDs), managed by the Ministry of Justice (MoJ).<sup>23</sup> The data includes rangatahi demographic information, offence types and dates, and the subsequent proceeding date and type (i.e., prosecution or the first type of diversionary action initiated by Police as a result of an investigation of the offence/s).<sup>24</sup> The table also includes information about the geographical location of each offence and the police station and district where rangatahi were dealt with for each offence. The table includes a total of 96,178 offence records and 57,851 distinct proceedings for 12,023 distinct rangatahi.

### *Other sources*

Police records of incidents and/or tasks<sup>25</sup> relating to the rangatahi above was sourced to construct a count of these as a matching variable for the evaluation. This data included 97,104 incidents/tasks for 10,858 rangatahi up to 31 May 2023.

Finally, data from Synergia was used to exclude rangatahi from the comparison pool who were referred to the pilot but did not enrol, since decline was unlikely to be

---

<sup>23</sup> The proceedings data in the YJNMDs is a subset of the Recorded Crime Offender Statistics (RCOS) data produced by Police. Data in the YJNMDs is restricted to information until the age of 25 for rangatahi and tamariki who offended between the ages of 10 and 17. At the time of receiving the data, it covered proceedings that occurred between January 2010 and June 2023, restricted to rangatahi who were SB participants or who recorded an AA diversion at some point during the pilot's lifetime.

<sup>24</sup> For more information about Police data definitions, see: [www.police.govt.nz/about-us/publication/national-recording-standard-november-2023](http://www.police.govt.nz/about-us/publication/national-recording-standard-november-2023).

<sup>25</sup> Incidents/tasks are police actions that do not relate to offending. The most common of these in the data provided were bail breaches, child protection reports, domestic disputes, and vehicle collisions.

random. In addition, this data is used to ensure (aggregated) referral and participation statistics are comparable with the official pilot statistics.

## Study populations

### *Treatment group (SB pilot participants)*

Of the 607 SB pilot participants, data limitations and the methodology used meant the treatment group for this evaluation includes a lower 471 participants who met the following selection criteria:

- having a NIA unique identifier and enrolment date in the SB participant table provided by Police (n = 603), and
- having police proceedings data in the YJNMDS linked to the NIA identifier (n = 579), and
- recording one or more proceedings in the police proceeding data less than six months before enrolment (n = 500), and
- enrolled to the pilot by 30 June 2022 (n = 483), and
- recording one or more offences in the proceeding that was closest to (and pre-dated) the enrolment date that were processed in Auckland (i.e., within the Counties Manukau, Auckland City, and Waitemata Police Districts) (n = 471).

In terms of the June 2022 restriction, this had to be applied so we had a minimum of 12 months of post-enrolment offending data to 30 June 2023 (12 months being a common timeframe used when analysing reoffending).

In terms of geography, the target population for the pilot was rangatahi residing (largely) in South Auckland suburbs. Therefore, both the treatment and matched comparison group (discussed in the next subsection) are restricted to rangatahi who were processed in Auckland for their offending. This is done since differences in institutional settings across regions, as well as time-region varying socioeconomic shocks (e.g., more frequent, and longer lock-down periods in Auckland during 2021 as a response to the COVID-19 outbreak) may impact rangatahi outcomes.<sup>26</sup> The data available for the evaluation did not include residential information, so we used the Police District where rangatahi were dealt with by Police for their offending as a proxy for residential location. This variable was found to be highly (and positively) correlated with residential location (see Appendix A for more information).

Table 3 presents the number of participants (treatment group), and mean value for a selection of demographic variables after applying the selection restrictions. The table also includes the same information for participants who were enrolled by June 2021, so had at least 24 months of reoffending data available, as well as the statistics obtained from the full cohort of participants for the same periods (as captured by Synergia). The table shows that in both the 12- and 24-month periods, the treatment group accounted for about 80% of all pilot participants. Relative to the Synergia data, the shares of rangatahi Māori (in the 12-month period) and females (in both periods)

---

<sup>26</sup> For more information about the rationale for restricting the study population to Auckland Police Districts, see sub-section 'Police District as a proxy for residential location' (and Table A3) within Appendix A.

are 2-4pp greater, while that of Pacific peoples is 1-2pp lower. Finally, the share of rangatahi with missing ethnicity information is 1-2pp greater in the treatment group.

*Table 3: Treatment group restrictions, sample sizes, and mean characteristics*

	Enrolled for at least			
	12 months		24 months	
	Synergia	Treatment group	Synergia	Treatment group
Participants	587	471	486	392
(Share)	100%	80%	100%	81%
Enrolment age	14.1	14.2	14.1	14.2
Females	29%	33%	30%	34%
<b>Ethnicity</b>				
Māori	70%	72%	71%	71%
Pacific Peoples	19%	17%	19%	18%
Other	10%	9%	10%	10%
Missing	1%	2%	0%	2%

**Source:** MoJ (2024), Police (2023), Synergia (2023). **Notes:** Enrolled for at least 12 month refers to enrolments by June 2022, and by June 2021 when at least 24 months.

Table 4 presents the (grouped) difference in days between enrolment, and the most recent prior police proceeding for the final treatment group (i.e., after applying all restrictions). The share of participants falls the longer the gap between the two dates becomes. For example, 31% of SB participants had a proceeding less than 30 days prior to enrolment, and over three quarters had a proceeding less than 90 days prior to enrolment. On average, the number of days between these two dates was 61 days (median of 51 days). This number of days seems to be reasonable since the average number of days between referral and enrolment for all 607 participants (using Synergia figures) was 44 days, suggesting a mean of 17 days between the proceeding date and referral to pilot.<sup>27</sup> In addition, similar to what was found when examining Synergia data, the average number of days between proceeding and enrolment increased by 12 days to 68 after March 2020 (first COVID-19 lockdown).

*Table 4: Number and share of final treatment group by (grouped) days between nearest prior proceeding and enrolment*

Days	Participants	Share	Cumulative share
0-30	147	31%	31%
31-60	119	25%	56%
61-90	99	21%	77%
91-120	53	11%	89%
121-182	53	11%	100%

**Source:** MoJ (2024), Police (2023).

### **Comparison pool (AA recipients)**

The initial comparison pool for the evaluation comprised 57,851 proceeding records for 12,203 distinct rangatahi who received AA at least once between September 2017 and August 2022 and were matched to the police proceedings data in the

<sup>27</sup> Note that this number is a rough approximation, since here, the gap was calculated between closest prior proceeding and enrolment. In some cases, the closest prior proceeding may not be the offending that led to referral to the SB pilot as some rangatahi reoffended and therefore recorded a proceeding after referral.

YJNMDS. Of those, 1,000 rangatahi and 5,489 proceedings were removed due to being duplicates, or since they were for rangatahi who were referred to or enrolled in the SB pilot. Next, the following restrictions were applied:

- Nearly 15,000 proceedings for 579 rangatahi were removed because the proceeding was before (after) the earliest (most recent) enrolment year/quarter recorded amongst all SB participants. In addition, a small number of records were removed since they captured information for rangatahi who were older (younger) than the oldest (youngest participant) in terms of either birth year, or age at reference date.
- Over 4,000 proceedings for about 36 rangatahi were removed because they recorded a greater number of offences than the SB participant who recorded the greatest number of offences during the 12-month period leading to the reference date. A similar restriction was used for prior incidents/tasks, resulting in the removal of over 2,000 incidents/tasks for more than 1,400 rangatahi.
- Finally, rangatahi who did not record any proceedings for an offence within an Auckland Police District were excluded. This resulted in removing nearly 29,000 proceedings for over 24,000 rangatahi.

Following the application of these restrictions, the final comparator pool includes 6,794 proceeding-rangatahi records for 2,394 rangatahi (Table 5).

*Table 5: Distinct rangatahi and proceedings counts by restriction, comparator pool*

Restriction	Rangatahi	Proceedings
All	12,203	57,851
AA cohort	11,203	52,362
Proceeding year/quarter	10,624	37,449
Age at reference date & birth year	10,618	37,436
Offences	10,582	33,246
Incidents	9,120	31,177
Offended in Auckland	2,394	6,794

**Source:** YJNMDS, MoJ (2024), Police (2023).

## Matching

Conceptually, to measure the impacts of the SB pilot on participants, their reoffending outcomes following enrolment need to be compared with their outcomes in an alternative scenario where they did not enrol (counterfactual, or potential outcomes). Then, averaging the individual-level differences across participants would reveal the average effect the pilot had on participants, or the Average Treatment Effect on the Treated (ATET).

Since only one of these scenarios is observable (i.e., participants' outcomes if enrolled), there is a 'missing data' problem. In this evaluation, this 'gap' is addressed by applying a Nearest Neighbour Matching (NNM) approach (Abadie & Imbens, 2002). NNM is a non-parametric matching technique that imputes the (missing) potential outcome for each participant by averaging the outcomes of the individual (or individuals) from a comparator group who had the greatest degree of 'similarity'. Determining which observation is most similar, is calculated by a weighted function that includes a set of observable characteristics. Here, similarity is measured via

Mahalanobis distance, with an additional set of observable characteristics that are required to be exactly matched (more on this in the next section).

Using this approach, the estimated impact of the pilot can then be written as:

$$E(\Delta\hat{Y}^{ATET}) = \frac{1}{N} \sum_{i=1}^N [Y_i - \hat{Y}_i(0)]$$

where  $E(\Delta\hat{Y}^{ATET})$  is the estimated mean change in outcome  $\hat{Y}$  amongst participants (ATET),  $N$  is the number of participants in the treatment group,  $Y_i$  is the (observed) participant level outcome, and  $\hat{Y}_i(0)$  is the participant level imputed potential (counterfactual) outcome.<sup>28</sup>

NNM was chosen over other commonly used algorithms (e.g., Propensity Score Matching), due to our preference to base similarity in terms of all observable characteristics, rather than a single matrix, such as outcome or enrolment probability.<sup>29</sup> However for reference, outcomes are re-estimated using other common methods such as Propensity Score Matching (PSM) and Inverse-Probability Weighted Regression Adjustment (IPWRA). Finally, outcomes are estimated amongst all participants (from the treatment group), and by a selection of specifications, including examining outcomes for Māori and non-Māori participants separately.<sup>30</sup>

In terms of causality, the primary assumption required for the NNM estimates to be unbiased is the *conditional independence* (CI) assumption. This requires that following matching, the only difference between the treatment and the matched control groups is that the first was treated (i.e., enrolled into the pilot).<sup>31</sup> That is, the treatment and matched control group are ‘as good as random’. While this cannot be tested formally,<sup>32</sup> this assumption is commonly explored by comparing the *balance* of each matching variable between the treatment and control groups. Here, the extent of similarity is summarised via the standardised (i.e., normalised) differences at each

<sup>28</sup> For more information about NNM ATET estimator, see Abadie et al. (2004).

<sup>29</sup> Note that if the treatment model is reasonably well specified, PSM will perform at least as well as NNM. The algorithm was executed using Stata’s ‘teffects nnmatch’ command. As discussed in Abadie and Imbens (2006, 2011), NNM estimators are not consistent when matching includes more than one continuous covariate. This is addressed by adding a bias-corrected estimator option (bisasdj). For more information about NNM (and other matching approaches), see Caliendo & Kopeinig (2005). Previous examples of using matching to evaluate social initiatives in New Zealand include Ku and de Boer (2018), Preval et al. (2021), and Webber (2023).

<sup>30</sup> Due to the relatively small number of participants who were Pacific peoples, this group was not examined separately.

<sup>31</sup> Note that since the teffect command (in Stata) is used for matching, only the conditional mean independence assumption (CMI) is needed to be met. This assumption means that after accounting for the matching variables, the treatment does not affect the conditional mean of each potential outcome. The CMI assumption is less restrictive since it allows for the conditional variance to depend on the treatment, while the CI assumption does not (StataCorp, 2019).

<sup>32</sup> Matching was selected as it was the most feasible approach within the context of the pilot. The lack of formal testing for this assumption limits the reliability NNM, or any other form of matching. For example, in the Maryland Scientific Method Scale (SMS), matching is ranked below (i.e., less reliable) than other commonly used methods such as Difference in Differences. For more information, see: <https://whatworksgrowth.org/resource-library/the-maryland-scientific-methods-scale-sms/>

level of matching variables.<sup>33</sup> To assess whether items are balanced, we use the common rule of thumb (since there is no formal test) that differences should be between -0.25 and 0.25.

### **Outcome variables**

To assess pilot effectiveness, we examine reoffending frequency and severity outcomes across 12- and 24-month windows from the reference date. For the control group, the reference date is their proceeding date/s, since it is the most likely date their Business as Usual (BAU) intervention commenced. For the same reason, the treatment group's reference date is their enrolment date.

Note that the data structure means that each SB participant appear only once, while non-participants from the comparator pool can appear more than once if they recorded more than a single proceeding between September 2017 and August 2022. In addition, due to the recency of the pilot, and the YJNMDS only containing offending data to 30 June 2023, outcomes are only measured over 12- and 24-month periods from reference date, or reoffending windows. In each of these windows, outcomes examine proxy reduction in reoffending frequency and severity.

The impact of the pilot on **reoffending frequency** is assessed by two outcome measures:

- whether rangatahi did not reoffend in each reoffending window since the reference date, measured via a dummy variable that equals one if no offences were recorded (zero otherwise)
- whether the total number of offences in each reoffending window was lower than the total number of offences in the equivalent period of months leading up to the reference date. For example, if person A recorded eight offences in the 12 months leading up to the reference date, but only four in the following 12 months, then the dummy variable will equal one (zero otherwise).

The impact of the pilot on **reoffending severity**<sup>34</sup> is assessed using two outcome measures:

- a dummy variable reflecting whether the most severe offence during the given reoffending window was at least 10% lower<sup>35</sup> than the most severe offence during the equivalent period leading up to that date
- a dummy variable reflecting whether the combined severity across all offences during the given reoffending window was at least 10% lower than that recorded in the equivalent period leading up to that date.

---

<sup>33</sup> Balance is measured as  $\Delta x = \frac{E(x_T) - E(x_C)}{\sqrt{\frac{S_T^2 + S_C^2}{2}}}$  with  $\Delta x$  being the standardised difference for matching variable  $x$ ,

$E(x_T)$  and  $E(x_C)$  are the treatment (T) and control (C) group's means (of  $x$ ), and  $S_T^2$  and  $S_C^2$  are the group's variance.

<sup>34</sup> See Appendix A for further information on how offence severity was assessed.

<sup>35</sup> A margin of error is used because seriousness scores are a statistical representation of how punitive courts are, on average, for each type of offence, so close scores of say 300 and 310 are unlikely to be meaningfully different.

In terms of outcome variable selection, part of the rationale for using individual level dummies is that the high-level goal of the youth justice sector is to reduce individual's offending severity and frequency, so we expect these variables to be closer proxies of this goal than other commonly used statistics, such as average changes in counts across an entire cohort of rangatahi. In addition, due to the relatively small size of the study population, cohort level measures are more likely to be affected by outliers (this was found to be the case in this analysis). However, for reference purposes, the evaluation also examines the impact of the pilot on the total number of offences committed, as well as the maximum, total, and mean reoffending severity.<sup>36</sup>

### **Matching variables**

The variables used for matching the treatment and control groups are summarised in Table 6, and are grouped by whether they capture time-invariant variables, pre-treatment information that occurred *by* the reference date (prior to and including), and pre-treatment information *at* the reference date.

In terms of time-invariant measures, matching variables include a female gender dummy, a Māori ethnicity dummy, a Pacific Peoples ethnicity dummy, and the (grouped) age at first recorded offence.<sup>37</sup> Next, variables that measure information *by* the reference date include the (grouped) maximum and total offence severity (using the MoJ offence seriousness score) in the 12 and 24 months leading up to the reference date, most severe offence ever, as well as the (grouped) number of offences recorded in these periods, and the (grouped) total incidents/tasks recorded.<sup>38</sup> In terms of interventions, this group also captures the most severe previous intervention (Warning, AA, FGC, Prosecution, and Other), and most serious prior care and protection involvement with Oranga Tamariki (None, Intake, Investigation, Intervention, Placement). Finally, variables measuring information at the reference date include the reference year/quarter, age (in years), and the proceeding outcome at that date.

Matching variables were largely selected based on whether they corresponded to (or were a close proxy of) YORST assessments items (e.g., total prior offences, highest level of previous Police and Oranga Tamariki interventions, age at first offence), since these were found to be valid predictors of youth reoffending in New Zealand (Mossman, 2016) and YORST scores were a key eligibility criterion when deciding whether to refer rangatahi to the pilot.

---

<sup>36</sup> Conceptually, since all rangatahi offended before the reference date, the outcome dummy indicators may capture natural changes in reoffending patterns that come with age. In the context of this study, we do not expect this to have a significant effect on the estimates due to the short period outcomes are measured for (i.e., up to 24-months post enrolment), and should be neutralised across groups via the matching process.

<sup>37</sup> Māori and Pacific Peoples ethnicity dummies were based on Total Response. More generally, dummy variables take the value of one if a condition is met, and zero otherwise. For example, the female gender dummy variable is equal to one if the participant is recorded as female, and zero if recorded as male, other gender, or unknown. The Age at first offence groupings are those used in the YORST assessment form (<10, 10-12, 13, 14, 15-17).

<sup>38</sup> In terms of offences and incidents counts, groupings were based on the distribution of the treatment group for each measure. For offending, counts were grouped to 0-1, 2-3, 4-9, 10-20, and 21-31 bins. For tasks/incidents, groupings were 0, 1, 2-3, 4-9, 10-20, and 21-34 bins.

Table 6: Matching variables

Group	Matching variables
Time invariant	Female gender, Māori ethnicity, Pacific Peoples ethnicity, grouped age at first recorded offence
By reference date	Maximum and total severity (using MoJ seriousness scores) in the prior 12 and 24 months (i.e., four variables), highest ever prior offence severity, grouped total number of offences in the prior 12 and 24 months (i.e., two variables), grouped total prior Police Incidents/Tasks, most severe prior Police intervention, most serious prior Care and Protection (C&P) involvement with Oranga Tamariki
At reference date	Reference year/quarter, age at reference date (in years), Proceeding outcome (most severe if more than one proceeding at reference date)

**Source:** MoJ (2024). **Notes:** For participants, proceeding outcome at reference date is measured for the proceeding closest to enrolment. The Māori and Pacific Peoples ethnicity dummies are based on total response.

In terms of the matching algorithm, the way variables are matched varies. Most variables that can be expressed numerically are defined as continuous in the matching process. These include the reference year/quarter, age at reference date, and all the prior offence severity measures. On the other hand, offences in the prior 12 months, offences in the prior 24 months, total prior police incidents/tasks, Pacific Peoples ethnicity, most severe prior police intervention, and most serious prior Oranga Tamariki care and protection intervention were all defined as categorical, meaning that participants were matched by their underlying levels. Finally, the female gender and Māori ethnicity dummies, as well as the (grouped) age at first offence, were *exactly* matched.<sup>39</sup>

### Assessing balance

Table 7 presents the balance table for the entire treatment group (SB pilot participants), and for the subgroup with 24 months of outcomes post-enrolment. For each reoffending window (12, 24), the table presents the mean value (when the variable is continuous) or distribution across levels (when categorical) for participants ( $E(T)$ ), matched control group ( $E(C)$ ), as well as the standardised differences across groups ( $\Delta x$ ). For reference, the table also includes the mean values for the wider comparator pool ( $E(Pool)$ ). Before discussing balance, the distribution across variables for participants is outlined, with some commentary regarding that of the entire comparator pool mentioned for reference purposes.

### Comparing participant characteristics to those of the wider comparison pool

Overall, any changes in participant values across windows are marginal (if any). In terms of prior offending, the table suggests that participants had a small number of offences recorded prior to enrolment. While this varies by the number of months prior, about 40% of participants recorded one offence, and over 80% up to three offences. Similarly, nearly 60% of participants recorded up to three Police incidents/tasks (ever). Conversely, about 20% of the rangatahi from the wider comparator pool recorded 10 or more prior offences (in the 12-month window), while this was the case for less than 2% of SB participants. In terms of offence severity, for participants both the score of the most severe offence (216-274), and total severity

<sup>39</sup> Ideally, we would have included a larger set of variables in the exact matching list (e.g., Pacific Peoples ethnicity, prior offending). However, insufficient numbers of exact matches prevented this.



(326-456) were lower than those recorded for the wider comparator pool (277-365, 727-1,002).

Table 7: Balance table, treatment group (SB participants) and matched control group

	Months from reference date (reoffending window)							
	12 months				24 months			
Matching variables	<i>E(Pool)</i>	<i>E(C)</i>	<i>E(T)</i>	$\Delta x$	<i>E(Pool)</i>	<i>E(C)</i>	<i>E(T)</i>	$\Delta x$
<b>Observations</b>	6,794	471	471	-	5,257	392	392	-
<b>Number of offences 12 months prior</b>								
1	27%	42%	40%	-0.04	28%	45%	43%	-0.04
2-3	26%	39%	41%	0.04	26%	39%	40%	0.03
4-9	28%	18%	18%	0.01	29%	15%	16%	0.02
10-20	16%	1%	1%	-0.02	14%	1%	1%	0.00
21+	4%	0%	0%	0.00	3%	0%	0%	0.00
<b>Number of offences 24 months prior</b>								
1	21%	38%	36%	-0.04	22%	40%	38%	-0.05
2-3	23%	36%	38%	0.04	23%	36%	38%	0.03
4-9	31%	24%	24%	0.00	31%	21%	22%	0.01
10-20	19%	2%	2%	0.00	18%	2%	2%	0.00
21+	7%	0%	1%	0.03	5%	0%	0%	0.07
<b>Prior offence severity (seriousness score)</b>								
Most severe offence 12m	287	184	216	0.13	277	182	216	0.13
Most severe offence 24m	332	196	257	0.21	318	197	253	0.20
Total offence severity 12m	795	312	356	0.09	727	275	326	0.13
Total offence severity 24m	1,002	354	456	0.18	921	325	428	0.20
Maximum prior severity	365	206	274	0.23	345	206	267	0.21
<b>Number of prior incidents/tasks</b>								
0	6%	4%	5%	0.02	7%	5%	5%	0.02
1	17%	21%	25%	0.11	17%	21%	26%	0.11
2-3	17%	29%	27%	-0.05	17%	27%	26%	-0.01
4-8	27%	32%	31%	-0.03	27%	33%	31%	-0.04
9-20	25%	12%	10%	-0.05	24%	13%	10%	-0.09
21+	9%	1%	1%	0.02	8%	1%	1%	0.03
<b>Age at first recorded offence</b>								
<10	4%	6%	6%	0.00	4%	6%	6%	0.00
10-12	25%	28%	28%	0.00	24%	29%	29%	0.00
13	20%	20%	20%	0.00	20%	19%	19%	0.00
14	20%	23%	23%	0.00	20%	22%	22%	0.00
15-17	31%	24%	24%	0.00	32%	24%	24%	0.00
<b>Most serious prior Police Intervention</b>								
Other	1%	3%	3%	0.01	1%	3%	4%	0.03
Warning	3%	3%	3%	0.01	3%	3%	3%	0.00
AA	50%	84%	83%	-0.02	50%	84%	84%	0.00
FGC	13%	4%	5%	0.05	12%	3%	4%	0.05
Prosecution	33%	6%	5%	-0.04	34%	7%	5%	-0.07
<b>Police Intervention for reference proceeding</b>								
Other	6%	4%	5%	0.03	6%	5%	6%	0.03
Warning	10%	5%	6%	0.05	10%	5%	6%	0.05

AA	51%	85%	84%	-0.01	52%	83%	84%	0.02
FGC	12%	4%	4%	0.00	11%	3%	4%	0.01
Prosecution	21%	2%	1%	-0.10	21%	4%	1%	-0.17
<b>Most serious prior Care and Protection (C&amp;P) involvement</b>								
None	49%	72%	70%	-0.06	49%	71%	69%	-0.03
Intake	2%	2%	2%	0.00	2%	2%	2%	0.04
Investigation	9%	12%	13%	0.02	9%	11%	12%	0.02
FGC	16%	7%	8%	0.02	16%	7%	8%	0.05
Placement	24%	7%	8%	0.06	24%	10%	9%	-0.04
<b>Demographic/other</b>								
Age at reference date	15.17	14.33	14.18	-0.09	15.13	14.33	14.16	-0.10
Female	23%	33%	33%	0.00	22%	34%	34%	0.00
Māori	54%	72%	72%	0.00	52%	71%	71%	0.00
Pacific Peoples	23%	22%	28%	0.14	23%	24%	29%	0.10

**Source:** MoJ (2024), Police (2023), Oranga Tamariki (2024).

Unsurprisingly, AA was the most common proceeding outcome both for the proceeding closest to enrolment, and across all proceedings prior to enrolment (83-84%). Only 5% of participants had recorded a prosecution by their reference proceeding, and only 1% at that date. This again, was different than the distribution of the wider comparator pool, who recorded far greater shares of prosecutions both by (33%) and at (21%) the reference proceeding.

Next, in terms of age when first coming to Police attention for their offending, nearly 30% of participants were aged 10-12, about one-fifth were aged 13, and nearly half were aged 14 or older. On the other hand, 6% were younger than 10. This was fairly similar to the distribution of the wider comparator pool, with the main difference being a greater share of rangatahi first offending between the ages of 15 and 17.

In terms of the most serious prior care and protection (C&P) involvement with Oranga Tamariki, 30% of participants recorded any prior C&P involvement. This share is lower than found in some other sources.<sup>40</sup> The 'highest' level of C&P involvement for 13% of participants was statutory assessment/investigation, followed by a C&P FGC referral or a C&P Placement (8% each). Participants infrequently recorded an Intake (following a Report of Concern) as their highest prior level of C&P involvement. This prior C&P share distribution for participants was considerably lower and less escalated than that seen for the wider comparator pool.

Finally, participants were aged 14 on average at the reference date, about one-third were female, and using total response ethnicity, over 70% identified as Māori and 28% as Pacific peoples. In comparison, rangatahi from the wider comparator pool were on average a year older, and included a lower share of females, Māori, and Pacific Peoples.

<sup>40</sup> For example, during the 2023 financial year, under 40% of the young people who had a youth justice FGC or were brought into custody had previously been in the statutory care system ([Oranga Tamariki BIM 2023](#)). In addition, only 44% of those born in New Zealand between 1993 and 2002 and recorded a YJ-FGC, also had prior C&P statutory involvement (Oranga Tamariki Evidence Centre, 2021).

## **Balance between the treatment and matched control groups**

When comparing the treatment group with the matched control group, the alignment of matching variables is significantly closer (as expected), where none of the standardised differences were large enough to be considered unbalanced (i.e., where  $\Delta x$  is outside of the range of -0.25 to 0.25).

While balanced, a few differences are relatively large and worth noting. First, participants recorded greater mean values in terms of all maximum and total prior severity measures. If greater severity is positively correlated with reoffending, then (controlling for all other factors) the estimated impact of the pilot may be *understated*.

In contrast, the matched control group was more heavily distributed towards a greater number of prior Police incidents/tasks (e.g., 2-3pp more with 9-20 incidents), and were (1-3pp) more likely to record a prosecution, especially at the reference proceeding. Again, if these are positively correlated with reoffending, then (controlling for all other factors) the estimated impacts of the pilot may be *overstated*.

Finally, the table suggests a greater share of Pacific Peoples amongst participants. This is somewhat expected, since most Pacific People in the Auckland region reside in South Auckland, and therefore would have been enrolled to the pilot (or at least referred) if they met the eligibility criteria. The impact this would have on the estimates is not clear, though in at least one example (Apatov, 2019), Pacific youth were more likely to record more favourable outcomes than comparable youth from other ethnic groups (i.e., after controlling for observable factors).

Overall, while there is no strong evidence that the treatment and matched control groups are unbalanced, it is not clear whether the outlined differences will have any impact on the estimates, and what magnitude or direction (i.e., positive or negative) any such biases would have. The potential contribution of these differences will be assessed in more detail as part of the analysis, where we examine the impacts for several sub-samples that decrease such differences across groups.

## **Limitations**

Despite our best efforts, the evaluation is not without limitations. There were issues with linking data and with the completeness of the police RCOS offending data. The Police “Person ID” can change when multiple identities for an individual in NIA get merged, which can affect data linkage on this variable. The offending data is not always complete as some rangatahi had no offences recorded, and some did not have any recorded within a reasonable timeframe before known interventions. These issues affected the characteristics and size of the treatment and matched control groups, which in turn, could bias the estimated impacts.<sup>41</sup>

Next, since the evaluation uses administrative data to construct the matching variables and outcomes, it is possible that these may not be sufficient proxies for the actual characteristics that affect reoffending outcomes, or the outcomes themselves. In addition, while the core aim of the pilot was to reduce reoffending frequency and

---

<sup>41</sup> If these errors occur randomly, the impact will be to bias the estimates towards zero. If non-random, the estimates may be positively or negatively biased (i.e., overstating or understating the impacts respectively).

severity, it is possible that the pilot affected other socio-economic outcomes (e.g., education, health) which cannot be measured given the available data.

Next, offending histories were only available up to 30 June 2023, so estimated reoffending outcomes over 24 months could only be estimated for rangatahi who enrolled sufficiently early to the pilot (i.e., up to 30 June 2021). Also, the YJNMDS data sourced for the evaluation only included rangatahi who were SB participants or who had at least one AA outcome recorded in NIA during the pilot's lifetime. Therefore, it excluded information about rangatahi in youth justice who had never received AA. Combined, these two limitations imply that any findings are only for those captured by the data and may not represent the full impact of the pilot.

In terms of matching, as discussed, the validity of the NNM estimates rely on the conditional independence assumption holding. While this assumption is explored by testing the *balance* between groups, it does not guarantee that the groups are balanced in terms of *unobserved* characteristics. Therefore, selection bias may still be present and bias the results, even after matching for observables.

Finally, a potentially significant difference between the treatment and matched control group is the choice of each group's reference date. Ideally, the reference date would be the intervention start date for both groups, but this was not identifiable in the data for the control group. For participants, the reference date is their SB pilot enrolment date, while for the matched control group it is taken to be their proceeding date. As a result of having this difference, the reoffending window differs between groups, since months may have passed between the proceeding and SB pilot enrolment dates. Discussions with Police suggested that for diversionary outcomes such as AA, the intervention (e.g., AA plan) is likely to start at or near the proceeding date but may be much later than the proceeding date for FGCs and prosecutions. Also of note is that enrolment to the pilot was reliant on not reoffending to a level that leads to FGC or prosecution during the enrolment process. This meant the treatment group excludes rangatahi who were referred to the pilot but were not enrolled due to serious reoffending. The comparator pool does not have this same process, possibly positively biasing the estimates (overstating benefits).

Combined, we suggest caution when attributing any estimated changes in participant reoffending outcomes solely to the pilot.

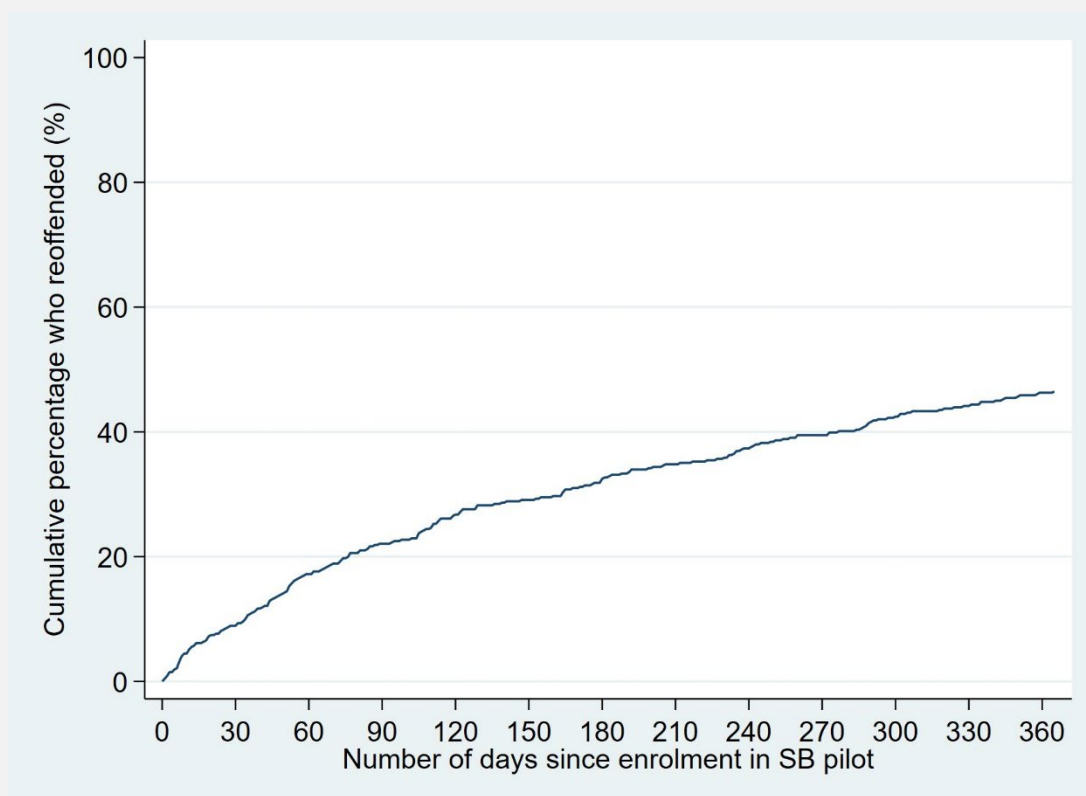
# Findings

## Reoffending by Social Bond pilot participants

This section focuses on 12-month reoffending outcomes for 471 rangatahi who participated in the SB pilot and who had enrolled prior to 30 June 2022 (so at least 12 months of reoffending data were available). A brief examination of 24-month reoffending rates is also presented for 392 rangatahi who had enrolled in the SB pilot prior to 30 June 2021.

Twenty-two percent of the 471 SB participants came to the attention of Police for a new offence within three months of enrolment in the pilot (Figure 2). Within six months, one-third of SB participants had reoffended, and within 12 months, 46% had reoffended. This means that the majority (54%) of SB participants did not reoffend within 12 months.

*Figure 2: Cumulative percentage of SB participants who had reoffended over the 12 months following enrolment*

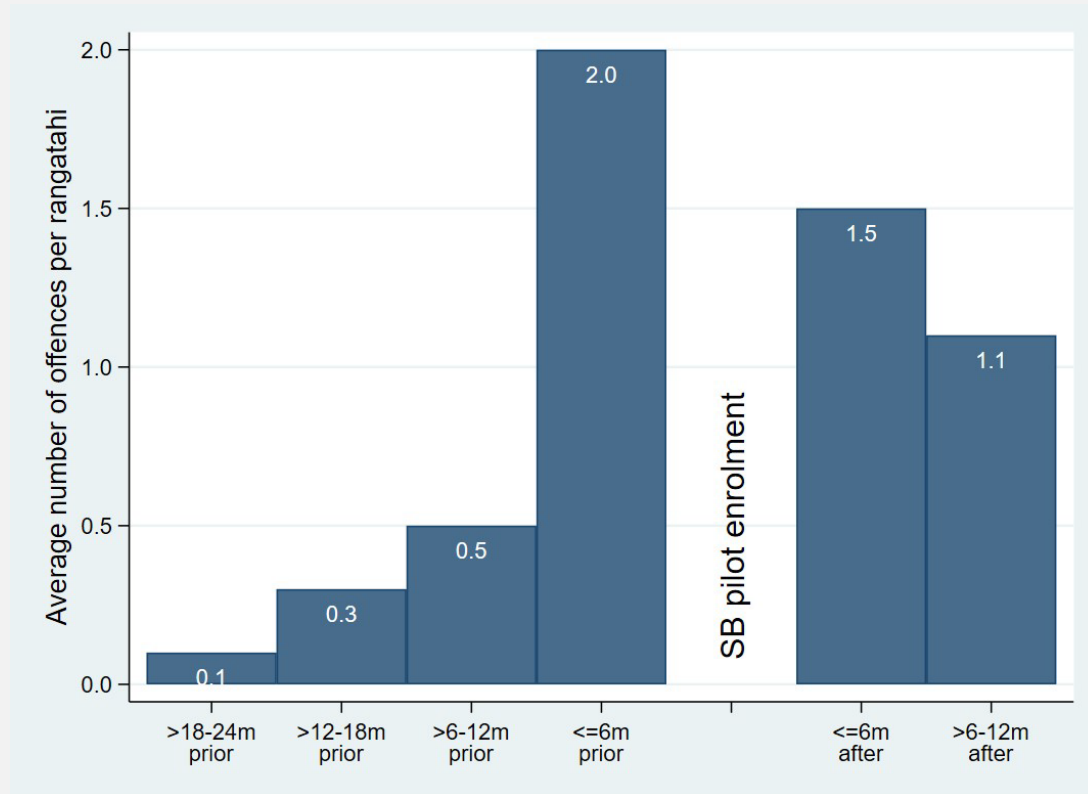


**Source:** MoJ (2024), Police (2023).

Most SB participants had not offended very much prior to the six months before enrolment in the pilot (Figure 3). There was an escalation in the average frequency and seriousness of offending in the six months leading up to enrolment. Over the 12 months after enrolment, the average number of offences committed by each rangatahi was 2.6, which was almost the same as the 2.5 offences on average in the 12 months prior.

The 471 SB participants committed a total of 70 more offences in the 12 months after enrolment than in the 12 months before (1,243 up slightly from 1,173). However, because the majority (54%) of SB participants did not reoffend, the smaller number of reoffenders committed more offences between them.

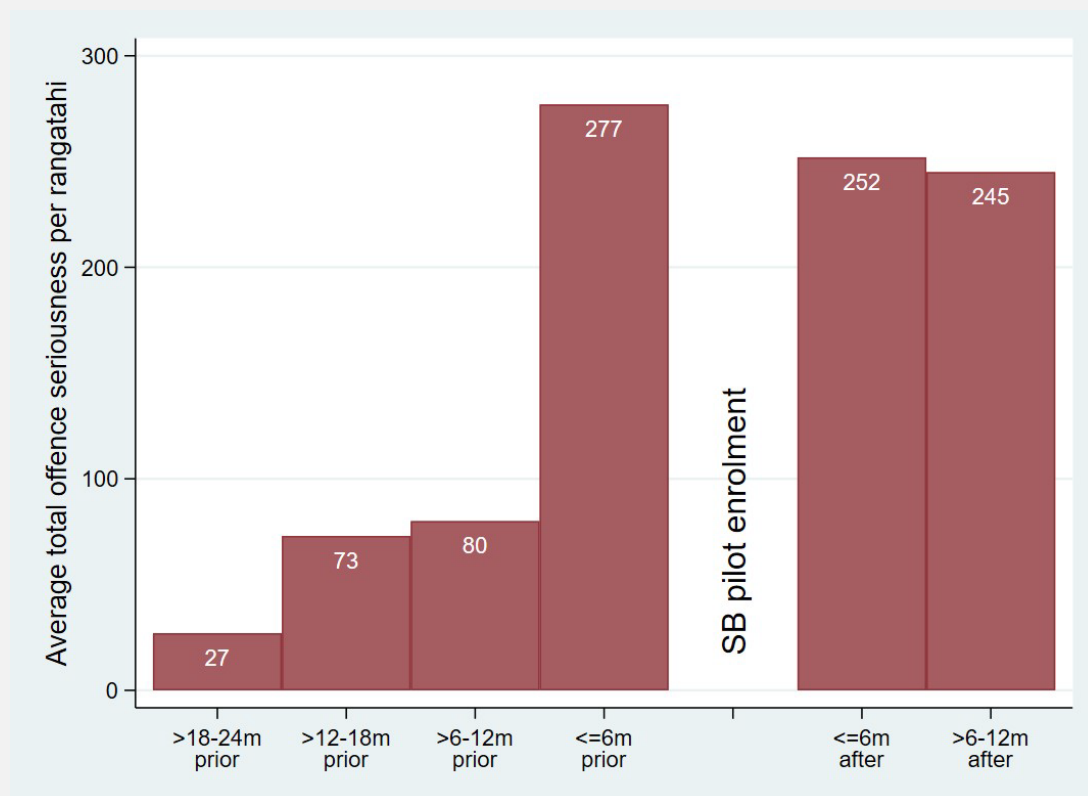
Figure 3: Average frequency of offences by SB participants in each six-month period before and after enrolment



Source: MoJ (2024), Police (2023).

Figure 4 shows that the average total seriousness of the offences committed by each rangatahi over the 12 months following enrolment (356) was 40% higher than in the 12 months before the orders (497).

Figure 4: Average total seriousness of offences by SB participants in each six-month period before and after enrolment



Source: MoJ (2024), Police (2023).

Table 8 shows that for all offence types, the number and percentage of SB participants who offended decreased or remained static in the 12 months after enrolment compared to the 12 months prior. However, the total number of offences of each type committed by SB participants increased in some cases.

Illegal use of a motor vehicle was the most common offence type committed by SB participants in the 12 months before enrolment (187 or 40% of participants committed 380 such offences). In the 12 months after enrolment, a much lower 120 (25%) of SB participants committed a greater number of such offences (498). A similar pattern was also seen for burglary i.e., fewer SB participants committing a greater total number of burglaries after enrolment compared to before.

The number and percentage of SB participants who committed an injury causing act or sexual offence, and the total number of such offences, all decreased in the 12 months after enrolment compared to the 12 months before. This was also seen for property damage, other theft offences, and public order or weapons-related offences.

The number and percentage of SB participants who committed a robbery-related offence or “all other offences” remained relatively static before and after enrolment. However, there was a small increase in the total number of robbery-related offences, and a larger increase in “all other offences” after enrolment.

*Table 8: Whether SB participants committed an offence of each type, and number of offences recorded by police, 12 months before and after enrolment*

<b>Offence type</b>	<b>No. and % who did not offend</b>	<b>No. and % who offended</b>	<b>Total offences</b>
<b>Injury causing acts or sexual</b>			
12 months prior	372 (79%)	99 (21%)	120
12 months after	420 (89%)	51 (11%)	100
<b>Robbery-related</b>			
12 months prior	434 (92%)	37 (8%)	44
12 months after	439 (93%)	32 (7%)	47
<b>Burglary</b>			
12 months prior	380 (81%)	91 (19%)	121
12 months after	409 (87%)	62 (13%)	164
<b>Illegal use of motor vehicle</b>			
12 months prior	284 (60%)	187 (40%)	380
12 months after	351 (75%)	120 (25%)	498
<b>Other theft-related</b>			
12 months prior	375 (80%)	96 (20%)	127
12 months after	424 (90%)	47 (10%)	77
<b>Public order or weapons</b>			
12 months prior	371 (79%)	100 (21%)	122
12 months after	400 (85%)	71 (15%)	86
<b>Property damage</b>			
12 months prior	352 (75%)	119 (25%)	157
12 months after	418 (89%)	53 (11%)	93
<b>All other offences</b>			
12 months prior	394 (84%)	77 (16%)	102
12 months after	395 (84%)	76 (16%)	178

**Source:** MoJ (2024), Police (2023).

Table 9 shows that the number and percentage of SB participants who offended within each seriousness band decreased in the 12 months after enrolment compared to the 12 months prior. However, the total number of offences in the three highest seriousness categories (between medium and high) increased. Much of the change seen in the 'Medium' seriousness category is due to the change described earlier for illegal use of motor vehicle offences. Similarly, much of the change seen in the 'Medium-High' and 'High' categories is due to the changes described earlier for burglary and robbery-related offences.



*Table 9: Whether SB participants committed an offence within each seriousness band, and number of offences recorded by police, 12 months before and after enrolment*

<b>Seriousness band</b>	<b>No. and % who did not offend</b>	<b>No. and % who offended</b>	<b>Total offences</b>
<b>Low</b>			
12 months prior	277 (59%)	194 (41%)	277
12 months after	363 (77%)	108 (23%)	191
<b>Low-Medium</b>			
12 months prior	356 (76%)	115 (24%)	139
12 months after	412 (87%)	59 (13%)	103
<b>Medium</b>			
12 months prior	197 (42%)	274 (58%)	546
12 months after	312 (66%)	159 (34%)	697
<b>Medium-High</b>			
12 months prior	379 (80%)	92 (20%)	111
12 months after	403 (86%)	68 (14%)	132
<b>High</b>			
12 months prior	392 (83%)	79 (17%)	100
12 months after	413 (88%)	58 (12%)	120

**Source:** MoJ (2024), Police (2023).

Table 10 presents four reoffending outcomes for SB participants. Over half (54%) of the participants did not reoffend in the 12 months after enrolling in the pilot. Around seven out of every 10 participants reduced the frequency and/or seriousness of their offending in the 12 months after enrolment compared to the 12 months before. There was little difference in reoffending outcomes for SB participants according to gender (Appendix Table B1). SB participants who were Pacific peoples or European generally had more positive reoffending outcomes than rangatahi Māori participants, although caution should be taken in interpreting differences because of small numbers (Appendix Table B2). Reoffending outcomes were more positive, the older rangatahi were at enrolment (Appendix Table B3). Finally, the table shows that in the 24-months after enrolment, 46% of rangatahi had not reoffended, and therefore 54% did reoffend. Around two-thirds of SB participants reduced the frequency and/or seriousness of their offending in the 24 months after enrolment compared to the 24 months before.

*Table 10: Reoffending outcomes for SB participants in the 12 and 24 months after enrolment*

<b>Outcome</b>	<b>Percentage of SB participants</b>	
	<b>12-months (n=471)</b>	<b>24-months (n=392)</b>
Did not reoffend	54%	46%
Reduced frequency of offending	69%	64%
Reduced total seriousness of offending	72%	66%
Reduced seriousness of most serious offence	70%	66%

**Source:** MoJ (2024), Police (2023). **Note:** Reduced frequency or seriousness of offending includes those who did not reoffend. For example, 69% of rangatahi reduced the frequency of their offending — comprising 54% who reduced their frequency from some non-zero number down to zero, and the remaining 15% who reoffended, but less often than before.

## Estimated impacts of the pilot

This section summarises the estimated impacts of the pilot. Table 11 presents the results for all SB participants (the treatment group), with each cell showing the percentage point (or pp) difference in mean outcome between the participants and the matched control group.<sup>42</sup>

In terms of reoffending frequency, column *None* indicates that SB participants were 10pp more likely to not reoffend in the 24 months following enrolment, or a 28% increase the share of participants who did not offend.<sup>43</sup> In addition, SB participants were also 8pp (14%) and 12pp (24%) more likely to record reduced reoffending frequency 12- and 24-months post enrolment, respectively. All these estimated impacts were statistically significant. While indicative of a positive change, the estimated impacts of the reoffending severity measures were not statistically significant.

*Table 11: Social bond pilot impact estimates by outcome measure and reoffending window*

Reoffending window	Frequency of reoffending		Severity of reoffending	
	None	Total	Max	Total
12 months (n=471)	0.014	0.084**	0.022	0.049
24 months (n=392)	0.101**	0.122**	0.024	0.061*

**Source:** MoJ (2024), Police (2023), Oranga Tamariki (2024). **Notes:** Differences measured as the Average Treatment Effect on the Treated. Stars indicate the level of statistical significance (\* = 10%; \*\* = 5%; \*\*\* = 1%).

Next, Table 12 presents the results solely for rangatahi Māori. The table finds no evidence for improvement in the likelihood of these rangatahi not reoffending, or in terms of reduced offence severity. However, as found overall, Māori participants were significantly more likely to reduce their reoffending frequency post-enrolment (an increase of 11pp, or a 21%). In terms of balance, Table B5 suggests that compared to their matched control group, Māori participants were unbalanced in terms of having greater (maximum) offence severity in the 24 months prior to the reference date, and at any time prior to their reference proceeding. In these variables, participants recorded greater mean values, possibly resulting in the estimates understating the true impacts of the pilot on this group.

*Table 12: Social bond pilot impact estimates for rangatahi Māori by outcome measure and reoffending window*

Reoffending window	Frequency of reoffending		Severity of reoffending	
	None	Total	Max	Total
12 months (n=337)	-0.026	0.070*	-0.007	-0.020
24 months (n=277)	0.074*	0.106**	-0.006	0.046

**Source:** MoJ (2024), Police (2023), Oranga Tamariki (2024). **Notes:** Differences measured as the Average Treatment Effect on the Treated. Stars indicate the level of statistical significance (\* = 10%; \*\* = 5%; \*\*\* = 1%).

For non-Māori participants, Table 13 suggests that in the 24 months following enrolment, the likelihood of not reoffending increased by 14pp (or 34%), as well as their likelihood to reduce their reoffending frequency (by 14pp, or 26%). In addition,

<sup>42</sup> Differences are measured as the Average Treatment Effect on the Treated (ATET).

<sup>43</sup> Percentage changes are calculated as the ratio of the ATET and imputed (missing) potential outcome (i.e., counterfactual). For more information and figures, see Table B4.

after 12 months, non-Māori were 11pp (or 13%) more likely to reduce the severity of their reoffending. All these estimated impacts were statistically significant. In terms of balance, there was no indication that any of the matching variables were unbalanced (Table B5).

*Table 13: Social bond pilot impact estimates for non-Māori by outcome measure and reoffending period*

Reoffending window	Frequency of reoffending		Severity of reoffending	
	None	Total	Max	Total
12 months (n=134)	0.045	0.074	0.072	0.114***
24 months (n=115)	0.142**	0.140**	0.047	0.083

**Source:** MoJ (2024), Police (2023), Oranga Tamariki (2024). **Notes:** Differences measured as the Average Treatment Effect on the Treated. Stars indicate the level of statistical significance (\* = 10%; \*\* = 5%; \*\*\* = 1%).

To explore the robustness of our main findings (i.e., Table), we re-estimated the impacts of the pilot using different sub-groups (excluding those prosecuted at reference proceeding, focusing on those who recorded AA outcome at reference proceeding, those who were processed in Counties Manukau, and clustering by prior offence severity), when using an alternative reference date to measure outcomes by/from, when (gradually) increasing the minimum number of nearest neighbour to four, and when applying alternative matching techniques (Propensity Score Matching, Inverse Probability Weighted Regression Adjustment).<sup>44</sup>

In this supplementary analysis, the most consistent finding was the reduction in participants' reoffending frequency in the 24-month following enrolment. As in our main findings, these improvements were large, positive, and with a similar magnitude across all specifications.

## Discussion

This evaluation examined the impacts of the SB pilot on participants' reoffending frequency and severity. The various specifications used repeatedly indicated that the SB pilot significantly reduced the reoffending frequency of participants, while findings regarding changes in the severity of offending were more mixed.

For example, amongst participants who had 24 months of outcomes data, the estimates suggest that relative to the counterfactual (i.e., being referred to a business-as-usual intervention), the share of participants who did not reoffend increased by 28%, and the share who reduced the frequency of their offending increased by 24%. On average, participants each recorded nearly two (1.976) fewer offences, seemingly driven by stronger reductions amongst those at the top of the offending distribution.

Aggregately, the pilot is estimated to have led to about 1,200 (1.976\*607) fewer offences being committed.<sup>45</sup> A crude approximation that applies the 2023Q4 NZD

<sup>44</sup> For the full discussion regarding the supplementary analysis, see Appendix C.

<sup>45</sup> Assuming that the mean reductions could be generalised across all participants.

average cost per offence (\$15,487),<sup>46</sup> indicates that the SB pilot may have saved over \$18.5m (1,200\*\$15,487) in avoided offending-related costs.

In terms of next steps, there would be value in re-estimating the impacts of the pilot in the future. This will have the advantage of a) estimating programme impacts when all participants have completed the programme and sufficient reoffending data is available, and b) estimating longer term impacts (i.e., beyond two years following enrolment).

In addition, since examining whether the SB pilot is more economically effective at addressing youth recidivism than the more traditional contractual approaches (or business-as-usual) is beyond the scope of this evaluation, we recommend that this question be specifically addressed by conducting a formal Cost Benefit Analysis (CBA). Such analysis would inform decision-makers whether the reduced reoffending found amongst participants was greater (in terms of monetised societal value) than the costs for operating the pilot.

As a minimum, such analysis should use this evaluation's results as input, while making the necessary adjustments to account for unobserved outcomes, impacts on specific offence types, and longer-term impacts. For a more comprehensive assessment, we suggest re-evaluating the impacts of the bond using roughly the same methodology in Stats NZ's Integrated Data Infrastructure (IDI) prior to the CBA, using new estimates as inputs.<sup>47</sup> Using the IDI as the data source will have the advantage of identifying benefits other than offending (e.g., education, health), the ability to include a broader range of matching variables than available in the YJNMDS, examining impacts over participants' family members, and using more recent data.<sup>48</sup>

Finally, the most comprehensive approach would include using the IDI when re-estimating impacts (i.e., so non-reoffending outcomes can also be included), but rather than using the same estimation strategy (i.e., matching), explore whether a more conceptually robust approach could be applied.<sup>49</sup>

---

<sup>46</sup> Adjusting the cost per criminal act as reported in Roper & Thompson (2006) to nominal values using the values reported by the New Zealand Treasury's CBAX tool in April 2024 (<https://www.treasury.govt.nz/information-and-services/state-sector-leadership/investment-management/investment-planning/treasurys-cbax-tool>).

<sup>47</sup> The IDI is a large research database that holds de-identified microdata about people and households' interactions with government agencies, NGOs, and survey responses. For more information, see: <https://www.stats.govt.nz/integrated-data/integrated-data-infrastructure/>

<sup>48</sup> The evaluation used the YJNMDS as the data source since key information was missing in the IDI during most of the project's lifetime. However more recently, this gap in information was addressed.

<sup>49</sup> For example, testing the possibility of applying a difference in differences approach. This is dependent on having the bond's data in the IDI, which was not added during most of the evaluation's lifetime (but has been added since).

# References

- Apatov, E. (2019). The Drivers Behind the Higher NEET Rate for Māori and Pacific youth: Insights from Administrative Data. Ministry of Business, Innovation & Employment - Hikina Whakatutuki.
- Abadie, A. and G. Imbens. (2002). Simple and bias-corrected matching estimators. Technical report. Department of Economics, University of California, Berkeley
- Abadie, A., Drukker, D., Herr, J. L., & Imbens, G. W. (2004). Implementing matching estimators for average treatment effects in Stata. *The Stata journal*, 4(3), 290-311.
- Abadie, A., & Imbens, G. W. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica*, 74(1): 235-267.
- Abadie, A., & Imbens, G. W. (2011). Bias-corrected matching estimators for average treatment effects. *Journal of Business & Economic Statistics*, 29(1): 1-11.
- Allen & Clarke (2023). Reducing Youth Reoffending in South Auckland Social Bond Pilot Evaluation: Qualitative outcomes evaluation. Allen & Clarke.
- Andres, J., & Dorsett, R. (2017). HMP Peterborough Social Impact Bond—Cohort 2 and final cohort impact evaluation. National Institute of Economic and Social Research.
- Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of economic surveys*, 22(1): 31-72.
- Fox, C., & Morris, S. (2021). Evaluating outcome-based payment programmes: Challenges for evidence-based policy. *Journal of Economic Policy Reform*, 24(1): 61-77.
- Gustafsson-Wright, E., Gardiner, S., & Putcha, V. (2015). The potential and limitations of impact bonds: Lessons from the first five years of experience worldwide. Global Economy and Development at Brookings.
- Jolliffe, D., & Hedderman, C. (2014). Peterborough Social Impact Bond: Final Report on Cohort 1 Analysis [Report prepared for the Ministry of Justice]. University of Leicester.
- King, G., & Nielsen, R. (2019). Why propensity scores should not be used for matching. *Political analysis*, 27(4), 435-454.
- Ku, B., & de Boer, M. (2018). Effectiveness of Driver Licensing Programmes Funded by the Ministry of Social Development. Wellington, New Zealand: Ministry of Social Development.
- Le, T., Riggs, L., & Hyslop, D. (unpublished) (2022). Evaluating the impact of the South Auckland Social Bond Pilot on reducing youth reoffending: A feasibility study. Report to Organa Tamariki. Motu Economics and Public Policy Research.

Malatest International and Oranga Tamariki Evidence Centre. (2021). Reducing Youth Offending Social Bond Pilot Process Evaluation. Wellington, New Zealand: Oranga Tamariki—Ministry for Children.

McLaren, K. (2010). Alternative Actions that Work: A review of the research on Police Warnings and Alternative. Action. Police Youth Services Group, New Zealand Police: Wellington.

MDRC. (2015). MDRC Statement on the Vera Institute's Study of the Adolescent Behavioral Learning Experience (ABLE) Program at Rikers Island.

Minister of Finance & Minister Responsible for Social Investment. (2017). The second NZ social bond: reducing reoffending in South Auckland. Cabinet Social Committee Paper.

Ministry of Health, on behalf of The New Zealand Government. (2015). Social Bonds 101: An Introduction to New Zealand's Social Bond pilot.

Ministry of Justice (2024). Youth Justice Indicators Summary Report. Youth Justice Indicators. Ministry of Justice.

Mossman, E. (2016). Research to Validate the New Zealand Police Youth Offending Risk Screening Tool (YORST) – Phase III. Commissioned by New Zealand Police.

Oranga Tamariki Evidence Centre (2018). Comparing Reoffending Rates for Participants of Police Alternative Action and Intention-to-Charge FGCs. [Unpublished research report].

Oranga Tamariki Evidence Centre (2021). Youth Justice Pathways: an examination of wellbeing indicators and outcomes for young people involved with youth justice. Wellington, New Zealand: Oranga Tamariki—Ministry for Children.

Parsons, J., Weiss, C., & Weiss, Q. (2016). Impact Evaluation of the Adolescent Behavioral Learning Experience (ABLE) Program. Vera Institute of Justice.

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.

Roper, T., & Thompson, A. (2006). Estimating the costs of crime in New Zealand in 2003/04 (No. 06/04). New Zealand treasury working paper.

Synergia (2023). Outcomes of the Genesis Social Bond Programme: Did Genesis Social Bond Programme Improve Outcome for Rangatahi and Youth People at Risk of Reoffending? Synergia Ltd.

Bakker, C (2023). Final in-depth review: Reduction in Youth Reoffending in South Auckland. Social Impact Bond. Acuo Limited, Wellington, New Zealand

Tan, S., Fraser, A., McHugh, N., & Warner, M. E. (2021). Widening perspectives on social impact bonds. *Journal of Economic Policy Reform*, 24(1): 1–10.

Ramsen, P (2016). Social impact bonds: state of play and lessons learnt. OECD Working Paper. OECD Local Economic and Employment Development.

StataCorp (2019). State Treatment-Effect Reference Manual: Potential Outcomes/Counterfactual Outcomes. Stata: Release 16. Pp. 221. College Station, TX: StataCorp LLC .URL: <https://www.stata.com/manuals16/te.pdf>

Wang, H., & Xu, X. (2022). Evidence-Based Analysis of Social Impact Bonds for Homelessness: A Scoping Review. *Frontiers in Psychology*, 13, 823390.

Webber, A (2023). Long-Term Impact of Teen Parent Units. *Te Atatū – Insights* September 2023. Wellington, New Zealand: Social Wellbeing Agency.

Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach*. Cengage learning.

# Appendix A

This appendix outlines the main data sources used in the evaluation, as well as several tests that informed the estimation strategy.

## Participants data – GYT (official)

Due to privacy limitations in the pilot’s consent form, data on SB pilot participants provided by GYT cannot be linked with other sources. However, the data can be examined on its own, and presented in this section for reference and comparison purposes with data from other sources.

Table A1 presents the demographic information for all SB participants, and by whether they recorded medium or high YORST score. Across groups, the average age at referral was 14.5, and most participants were male (with greater share amongst those with medium YORST) and Māori (with greater share amongst those with high YORST). Pacific Peoples were the second largest ethnic group and were more likely to record a medium YORST. Finally, the high YORST group recorded on average a greater YORST score (by definition) and were also more likely to exit the pilot prematurely (26% compared with 22% in the medium YORST group).<sup>50</sup>

*Table A1: Participant characteristics*

	All	Medium YORST (40-59)	High YORST (60-100)
Participant number	607	409	198
Age at referral	14.5	14.5	14.5
female	30%	28%	33%
<b>Prioritised ethnicity</b>			
Māori	70%	67%	76%
Pacific Peoples	19%	21%	14%
All other	10%	10%	9%
Missing	2%	2%	1%
Premature exit	23%	22%	26%
YORST at referral	55	49	67
YORST 60+	33%	0%	100%

Source: Synergia (2023).

## Participants data – Police

New Zealand Police (Police) provided the unique personal identifier from the National Intelligence Application (NIA) for 603 SB pilot enrollees (out of 607 participants) who had an ‘alert’ in NIA of “Genesis Social Bond Client”, along with the alert creation date (which was usually the pilot enrolment date). This data did not include rangatahi who were referred to the pilot but were not enrolled.

<sup>50</sup> For more information about the characteristics and outcomes of these participants, see Allen & Clarke (2023).



## Police Proceedings data – Ministry of Justice (YJNMDS)

Ministry of Justice provided extract of Police proceeding data from the Youth Justice National Minimum Dataset (YJNMDS) under a cross agency MoU. This information was extracted specifically to SB participants and non-participants that recorded a diversion to Alternative Action (AA) during the pilot period. The proceedings data captures history of offending<sup>51</sup> from January 2010<sup>52</sup> for individuals who are recorded in NIA with a role of ‘Offender’ in relation to an offence (other roles in an offence can be Victim, Witness, Suspect etc.). The data included proceedings to 30 June 2023.

The police proceedings table includes rangatahi unique identifiers (NIA Person IDs), demographic information, offence, and occurrence<sup>53</sup> details (offence and proceeding dates, offence type, location, and severity), and details from the resulting police proceeding (e.g., dates, outcome). The table includes a total of 97,090 records regarding 82,415 distinct offences (i.e., the same offence could be conducted by more than one rangatahi), 61,904 distinct occurrences, and 48,109 distinct proceedings for 12,817 distinct rangatahi. In terms of demographics, 73% of rangatahi were male, over half (55%) identified as Māori, 11% as Pacific Peoples, 48% as all other (known) ethnicities, and 3% with unknown ethnicity (using multiple response ethnicity). The mean birth year was 2004, with more than half born between 2001 and 2004, and nearly 90% between 2001 and 2007.

On average, rangatahi recorded eight offences, six occurrences, and five proceedings. However, these means are affected by the uneven distribution of justice related activity. In more detail, Table A2 presents the distribution of the records for rangatahi by the number of offences, occurrences, and proceedings in the proceeding table. In all measures, the share of rangatahi falls as frequency increases, with one third of rangatahi recording only a single proceeding. As expected, the distribution of rangatahi by the number of proceedings recorded was more heavily concentrated in a smaller number of records (than occurrences or offences), since a single proceeding can include more than one occurrence and/or offence.<sup>54</sup>

*Table A2: Frequency of offences, occurrences, and proceedings*

Records	Offences	Occurrences	Proceedings
1	28%	32%	33%
2	15%	16%	18%
3-4	17%	18%	19%
5-10	20%	18%	18%
11-20	10%	9%	8%
21+	10%	7%	4%

Source: MoJ (2024).

<sup>51</sup> Offences are recorded in NIA when there is *prima facie* evidence that on the balance of probability the matter amounts to an offence defined in New Zealand law. Most offences will be admitted by the young person and/or be formally proven in court.

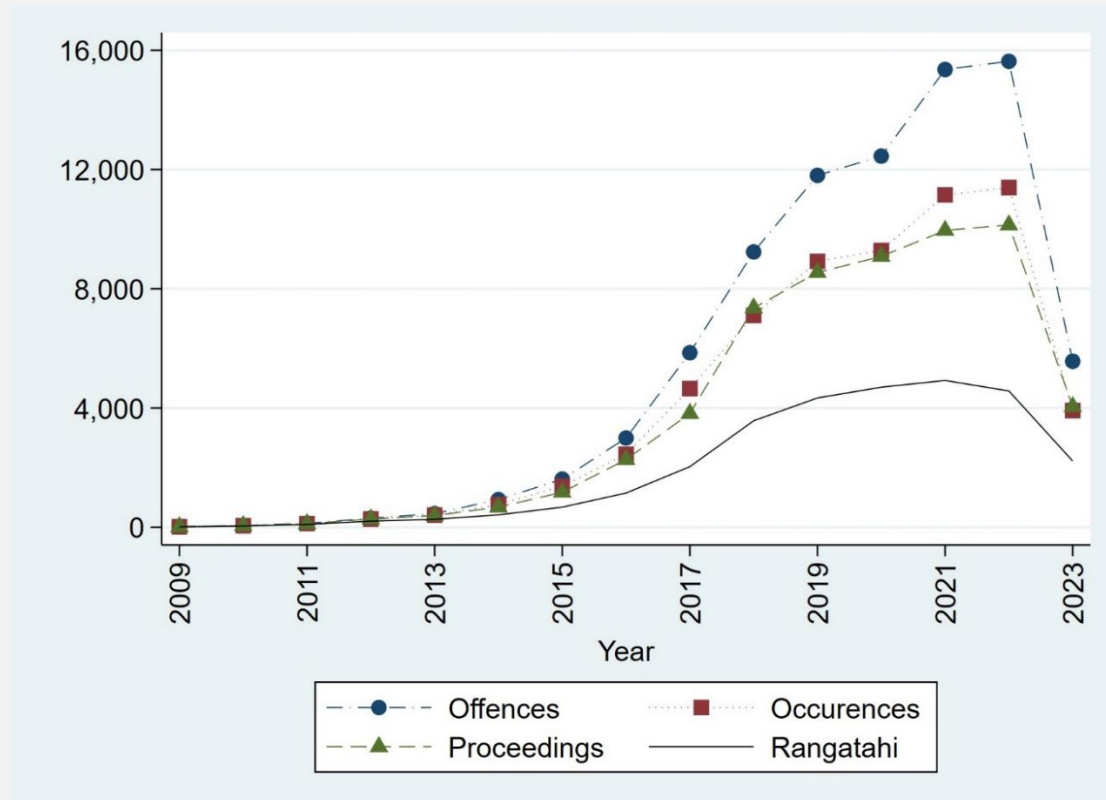
<sup>52</sup> The data includes offences where the date of proceeding is on or after 1 January 2010, but some of the offences were committed before this date. The data includes offences up to the offender turning 25.

<sup>53</sup> Multiple offences by an individual are grouped by police as a single ‘occurrence’ when they occur at the same time and location.

<sup>54</sup> Similarly, a single occurrence can include more than one offence.

Figure A1 presents the number of offences, and distinct occurrences, proceedings, and rangatahi in each year between 2009 and 2023. The figure shows increase over in all measures, with all but rangatahi count peak in 2022 (peak in 2021). The lower values in earlier years reflect that the data was only captured for rangatahi who were diverted to Alternative Action outcome during the pilot's period (September 2017 to August 2022), hence many rangatahi in the sample were too young to have a record, while the lower counts in 2023 reflect that the data only extended until June of that year (extraction date).

Figure A1: Annual distinct offences, occurrences, proceedings, and individuals, 2009-2023



**Source:** MoJ (2024). **Notes:** offences and occurrences are aggregated by occurrence year (using occurrence last possible date). Proceedings and rangatahi aggregated by year of proceeding.

Finally, 41% of proceedings resulted in an AA diversion, followed by 28% with prosecutions, 14% with warnings, 10% with referrals to a family group conference (FGC), and 6% with other outcomes (e.g., No further action).<sup>55</sup>

### Police District as a proxy for residential location

Since rangatahi residential location is not included in the YJNMDS, this section examines whether using the Police District the offence was processed within could be used as a proxy. This is key for the evaluation since the pilot was geographically bounded to *South Auckland*, and therefore their comparison group should ideally also be this area so to control for geographical differences that may affect outcomes.

<sup>55</sup> This distribution prioritises the greatest level of intervention if rangatahi recorded more than one outcome for the same proceeding (2.7% of records). This occurs when different proceeding-related offences require are responded to by different types of interventions. Without this prioritisation, the share of Prosecutions (FGS) falls (increases) by 1pp.

In terms of overall offence distribution, 29% of all offences from the proceeding table occurred within the various Auckland Police Districts,<sup>56</sup> and followed by the Bay of Plenty (14%), Central North Island (12%), and with 8% of all offences in each of the Waikato, Wellington, and the Canterbury districts. Within Auckland, 41% of offences were committed in the Counties-Manukau Police District, 30% in Waitematā, and 29% in Auckland City. Overall, the Counties-Manukau Police District accounts for 12% of all offences nationally.

To examine the relationship between Police District and residential location, the proceeding table was linked with a source (Police) that includes 18,236 AA diversions (i.e., proceedings where AA was the response used by police) for 12,624 distinct rangatahi, where for some, residential address at the time was included. Matching this source with the proceeding table by rangatahi NIA identifier and restricting to instances where the proceeding and AA diversion occurred in the same date, suggests 92% of the offences processed within the Auckland Police Districts were committed by residents of this region. Next, when linking the sources and examining the proceeding that was closest to the AA diversion date (and predated it by no longer than 6 months), the share of offences processed within Auckland that were committed by Auckland residents increased to 94%. For the participants from the treatment group, this share is greater, at 98%. Similarly, using all available offending records from the Integrated Data Infrastructure (IDI) Police table (restricting for youth that received a non-court action outcome, a proxy for Alternative Action), found that 95% of all offences occurring in the Auckland region were committed by Auckland residents. Therefore, there the likelihood that an offence processed within Auckland was committed by a resident of Auckland is very high.

In terms of selecting a comparator pool, the analysis targeted rangatahi who were processed within the three Auckland Police Districts (Counties Manukau, Auckland City, and Waitemata). As discussed in the report, the rationale for restricting rangatahi to Auckland was to avoid any biases emerging from either permanent, or time invariant differences across regions that may affect rangatahi outcomes, since participants were required to reside in Auckland, largely in South Auckland suburbs.

To explore this more formally, Table A3 summarises in the relationship between all matching variables used in the evaluation with all 12-month outcomes for all rangatahi who recorded an Alternative Action during the pilot period. The table present the marginal effects of the changes in variables and levels with each outcome. In terms of location, the table consistently shows that rangatahi processed in the Auckland District were on average less likely to record reduced reoffending frequency and severity. On the other hand, no such differences were found between those in Counties Manukau, and the rest of the region. Similarly, when analysing Auckland rangatahi (table not included), once again not statistically significant (and very small) differences were found between rangatahi processed in Counties Manukau, and elsewhere in the region. Overall, the finding support the evaluation's decision to only include Auckland rangatahi for the comparator pool, and not to restrict to specific Police districts within the region.

---

<sup>56</sup> Using Police District categories. This combined the Auckland City (6% of all offences), Counties-Manukau (12%), and Waitematā (9%) districts.

Table A3: Relationship between matching variables and 12-month outcomes, all rangatahi with AA diversion during pilot period

Matching variables	Frequency of reoffending		Severity of reoffending	
	None	Total	Max	Total
<b>Number of offences 12 months prior (base group: 0-1 offences)</b>				
2-3	-0.071***	0.151***	0.041***	0.075***
4-9	-0.124***	0.246***	0.061***	0.117***
10-20	-0.176***	0.332***	0.041	0.116***
21+	-0.166***	0.375***	0.049	0.151***
<b>Number of offences 24 months prior (base group: 0-1 offences)</b>				
2-3	-0.022*	-0.045***	-0.041***	-0.047***
4-9	-0.065***	-0.063***	-0.085***	-0.083***
10-20	-0.118***	-0.026	-0.116***	-0.104***
21+	-0.154***	-0.010	-0.162***	-0.146***
<b>Prior offence severity (seriousness score)</b>				
Most severe offence 12m	0.006**	0.011***	0.035***	0.020***
Most severe offence 24m	-0.006***	-0.006**	-0.008***	-0.007**
Total offence severity 12m	-0.001	-0.000	-0.001	0.004**
Total offence severity 24m	0.002***	0.001	0.001	0.001
Maximum prior severity	0.004**	0.002	-0.004**	-0.002
<b>Number of prior incidents/tasks (base group: no prior incidents)</b>				
1	0.120***	0.100***	0.068***	0.051***
2-3	0.009	-0.012	-0.009	-0.028**
4-8	-0.023*	-0.024*	-0.019	-0.042***
9-20	-0.048***	-0.059***	-0.038**	-0.071***
21+	-0.042**	-0.041*	-0.015	-0.047**
<b>Age at first recorded offence (base group: 15-17 years old)</b>				
<10	-0.073***	-0.056***	-0.025	-0.021
10-12	-0.032***	-0.019	-0.010	-0.012
13	-0.031***	-0.007	-0.017	-0.004
14	-0.004	0.012	-0.007	-0.011
<b>Most serious prior Police Intervention (base group: 'All other' outcomes)</b>				
Warning	0.034	0.016	0.022	0.021
AA	0.247***	0.220***	0.195***	0.193***
FGC	0.269***	0.283***	0.239***	0.243***
Prosecution	0.289***	0.319***	0.292***	0.301***
<b>Police Intervention for reference proceeding (base group: 'All other' outcomes)</b>				
Warning	-0.059***	-0.031**	-0.025*	-0.031**
AA	0.001	0.021	0.018	0.017
FGC	-0.044***	-0.039**	-0.019	-0.031*
Prosecution	-0.085***	-0.062***	-0.062***	-0.060***
<b>Most serious prior Care and Protection (C&amp;P) involvement (base group: None)</b>				
Intake	-0.324***	-0.352***	-0.287***	-0.290***
Investigation	-0.302***	-0.374***	-0.278***	-0.320***
FGC	-0.222***	-0.267***	-0.192***	-0.206***
Placement	-0.234***	-0.289***	-0.213***	-0.236***
<b>Demographic/other</b>				
Age at reference date	0.010***	0.023***	0.016***	0.020***
Female	0.056***	0.085***	0.048***	0.059***
Māori	-0.019***	-0.032***	-0.019**	-0.024***
Pacific Peoples	-0.012	0.001	0.007	-0.000

Auckland Police Districts	-0.018**	-0.036***	-0.039***	-0.043***
Counties Manukau District	-0.002	-0.008	-0.015	-0.005
Intercept	0.157	-0.281	0.911***	1.080***
Observations	33,218	33,218	33,218	33,218

**Source:** MoJ (2024), Police (2023), Oranga Tamariki (2024). **Notes:** Results from Ordinary Least Square specifications. Person identifier/proceeding date units of observations. Standard errors were clustered at the personal identifier level. Stars indicate the level of statistical significance (\* = 10%; \*\* = 5%; \*\*\* = 1%). Reference year/quarter was included in the analysis, as a continuous measure with a unique, and linearly increasing value with more recent data (e.g., 2020q1 = 230, 2020q2 = 231). The value of this measure was small and statistically insignificant when measuring reoffending frequency. For the reoffending severity measures, each increase in quarter was associated with a statistically significant -0.002 and -0.003 in max and total severity, respectively.

## Assessing offence severity

Offence severity (seriousness) was assessed using the *Justice Sector Seriousness Scale* (August 2022 version). The scale, developed by the Ministry of Justice, is a way of comparing the seriousness of different types of offences based on actual court sentencing data.

Seriousness scores are a statistical representation of the average number of days of imprisonment imposed by the courts for each offence. Statistical equivalences are used for people who receive community-based sentences or fines. Murder has the highest seriousness score at just under 12,000.

A seriousness score of 200 for an offence implies that, on average, the courts impose sentences that are twice as severe as those imposed on an offence with a score of 100. However, caution is required with this interpretation as the way in which scores are calculated means there is sizable variance in some averages, particularly for offences of ‘moderate’ seriousness where the types and lengths of sentences imposed can vary considerably.

Because of the rather convoluted methodology, it is not possible to calculate error variances for the seriousness scores. However, because it needs to be acknowledged that seriousness scores, and therefore summary values calculated from them, have error variance, it was decided to use a 5% margin of error around the total seriousness scores when comparing the before and after periods. In effect this means that total seriousness scores for an individual need to differ between the two periods by more than 10% to be considered different. A 5% margin of error was also used when comparing the highest seriousness scores between the before and after periods.

The seriousness scale is developed from District and High Court sentencing data relating to adults. While there could be debate over the applicability of the scale to offending by young people, it is still considered useful to assess relative changes in the seriousness of offending by young people. Developing a similar scale based on the diverse range of ways in which cases are dealt with in the youth justice system would be extremely difficult.

In terms of reduced reoffending severity, if a rangatahi committed the offences shown in Table A3, the total seriousness of offending dropped from 593.9 in the 12 months before enrolment, to 112.8 in the 12 months after enrolment. Given the difference in these figures is much greater than 10%, this rangatahi is counted as having reduced his or her total seriousness of offending. The rangatahi is also

counted as having reduced the seriousness of the most serious offence, as the highest score dropped from 353.6 to 95.0, and this difference is considerably more than 10%. Note that this rangatahi reduced the frequency of his or her offending from four to two offences.

In the table, if the total seriousness of offending in the 12 months after enrolment had been much higher at say 570.0, given that this figure is only 4% lower than the figure of 593.9 in the 12 months before enrolment, the total seriousness would be taken to be the same (i.e., not statistically different) in the two periods.

*Table A3: Fictitious example of seriousness scores for a rangatahi*

12 months before enrolment		12 months after enrolment	
Offences committed in the period	Seriousness scores	Offences committed in the period	Seriousness scores
Burglary (\$500 to \$5,000) by night	353.6	Unlawfully gets into motor vehicle	95.0
Unlawfully takes motor vehicle	197.8	Common assault (manually)	17.8
Shoplifts (value under \$500)	26.5		
Wilful damage	16.0		
<b>Total</b>	<b>593.9</b>	<b>Total</b>	<b>112.8</b>

**Source:** MoJ (2024).

The seriousness bands used in Table 9 earlier in the report were:

- Low (0 to 23)
- Low-Medium (>23 to 43)
- Medium (>43 to 201)
- Medium-High (>201 to 355)
- High (>355)

# Appendix B

*Table B1: Reoffending outcomes for SB participants in the 12 months after enrolment, by gender*

<b>Outcome</b>	<b>Male</b> (n=316)	<b>Female</b> (n=155)
Did not reoffend	53%	55%
Reduced frequency of offending	68%	70%
Reduced total seriousness of offending	72%	72%
Reduced seriousness of most serious offence	71%	69%

**Source:** MoJ (2024), Police (2023).

*Table B2: Reoffending outcomes for SB participants in the 12 months after enrolment, by ethnicity (prioritised)*

<b>Outcome</b>	<b>Māori</b> (n=341)	<b>Pacific peoples</b> (n=83)	<b>European</b> (n=44)
Did not reoffend	50%	59%	68%
Reduced frequency of offending	66%	76%	70%
Reduced total seriousness of offending	68%	82%	75%
Reduced seriousness of most serious offence	67%	78%	75%

**Source:** MoJ (2024), Police (2023).

*Table B3: Reoffending outcomes for SB participants in the 12 months after enrolment, by age at enrolment*

<b>Outcome</b>	<b>10-13 years</b> (n=156)	<b>14-15 years</b> (n=215)	<b>16-18 years</b> (n=100)
Did not reoffend	46%	54%	63%
Reduced frequency of offending	62%	70%	77%
Reduced total seriousness of offending	65%	71%	84%
Reduced seriousness of most serious offence	66%	68%	81%

**Source:** MoJ (2024), Police (2023).

Table B4: Mean and differences in outcomes by specification

	12 months reoffending window				24 months reoffending window			
	Frequency of reoffending		Severity of reoffending		Frequency of reoffending		Severity of reoffending	
	None	Total	Max	Total	None	Total	Max	Total
<b>All participants</b>								
Imputed mean outcome	0.5209	0.5697	0.6791	0.6666	0.3629	0.5130	0.6317	0.5976
Observed mean outcome	0.5350	0.6539	0.7006	0.7155	0.4643	0.6352	0.6556	0.6582
Difference (pp)	0.0141	0.0842	0.0216	0.0489	0.1014	0.1222	0.0239	0.0606
Difference (%)	0.027	0.148	0.032	0.073	0.279	0.238	0.038	0.101
<b>Māori participants</b>								
Imputed mean outcome	0.5249	0.5595	0.6743	0.6600	0.3523	0.5146	0.6485	0.5927
Observed mean outcome	0.4985	0.6291	0.6677	0.6795	0.4260	0.6209	0.6426	0.6390
Difference (pp)	-0.0264	0.0695	-0.0067	0.0195	0.0737	0.1064	-0.0059	0.0463
Difference (%)	-0.050	0.124	-0.010	0.030	0.209	0.207	-0.009	0.078
<b>Non-Māori participants</b>								
Imputed mean outcome	0.5820	0.6425	0.7113	0.6922	0.4144	0.5294	0.6399	0.6210
Observed mean outcome	0.6269	0.7164	0.7836	0.8060	0.5565	0.6696	0.6870	0.7043
Difference (pp)	0.0449	0.0739	0.0723	0.1137	0.1421	0.1402	0.0471	0.0834
Difference (%)	0.077	0.115	0.102	0.164	0.343	0.265	0.074	0.134
<b>Not prosecuted at reference proceeding</b>								
Imputed mean outcome	0.5188	0.5687	0.6669	0.6527	0.3791	0.5195	0.6399	0.6008
Observed mean outcome	0.5322	0.6502	0.6974	0.7124	0.4625	0.6305	0.6512	0.6537
Difference (pp)	0.0133	0.0815	0.0305	0.0598	0.0834	0.1110	0.0113	0.0529
Difference (%)	0.026	0.143	0.046	0.092	0.220	0.214	0.018	0.088
<b>Diverted to AA in reference proceeding</b>								
Imputed mean outcome	0.5609	0.5837	0.7060	0.6686	0.4253	0.5390	0.6697	0.6279
Observed mean outcome	0.5567	0.6574	0.7128	0.7305	0.4817	0.6311	0.6646	0.6707
Difference (pp)	-0.0043	0.0737	0.0068	0.0619	0.0565	0.0921	-0.0051	0.0429
Difference (%)	-0.008	0.126	0.010	0.093	0.133	0.171	-0.008	0.068



**Processed in South Auckland (Counties Manukau Police District)**

Imputed mean outcome	0.5849	0.6436	0.7597	0.7250	0.4136	0.5588	0.6666	0.6420
Observed mean outcome	0.5387	0.6749	0.7028	0.7183	0.4621	0.6515	0.6515	0.6553
Difference (pp)	-0.0462	0.0313	-0.0569	-0.0067	0.0486	0.0927	-0.0151	0.0133
Difference (%)	-0.079	0.049	-0.075	-0.009	0.117	0.166	-0.023	0.021

**Max lifetime offence severity less or equal to participants' median (198)**

Imputed mean outcome	0.5417	0.5895	0.6098	0.6388	0.4228	0.4925	0.4943	0.5417
Observed mean outcome	0.5414	0.6504	0.6353	0.6805	0.4570	0.6018	0.5611	0.5414
Difference (pp)	-0.0004	0.0608	0.0255	0.0417	0.0342	0.1093	0.0668	-0.0004
Difference (%)	-0.001	0.103	0.042	0.065	0.081	0.222	0.135	-0.001

**Max lifetime offence severity greater than participants' median (198)**

Imputed mean outcome	0.4523	0.5816	0.7286	0.6922	0.3840	0.5727	0.7535	0.4523
Observed mean outcome	0.5268	0.6585	0.7854	0.7610	0.4737	0.6784	0.7778	0.5268
Difference (pp)	0.0745	0.0770	0.0567	0.0688	0.0897	0.1057	0.0243	0.0745
Difference (%)	0.165	0.132	0.078	0.099	0.234	0.185	0.032	0.165

**Offended 12/24 months post reference date**

Imputed mean outcome	-	0.2001	0.2796	0.2752	-	0.2172	0.4280	0.3603
Observed mean outcome	-	0.3059	0.3699	0.4018	-	0.3190	0.3571	0.3619
Difference (pp)	-	0.1059	0.0903	0.1266	-	0.1018	-0.0708	0.0016
Difference (%)	-	0.529	0.323	0.460	-	0.469	-0.166	0.004

**Source:** MoJ (2024), Police (2023), Oranga Tamariki (2024). **Notes:** The table presents the mean observed and imputed outcomes for participants from the various matching specifications. Differences are presented in terms of percentage point and percent from the imputed value.

Table B5: Balance table, various specifications

Matching variables/months	Months from reference date (reoffending window)															
	Māori		Non-Māori		Non-prosecution		AA		South Auckland		Lower severity		Greater severity		Offended	
	12	24	12	24	12	24	12	24	12	24	12	24	12	24	12	24
<b>Number of offences 12 months prior</b>																
0-1	-0.02	-0.04	-0.10	-0.05	-0.04	-0.07	-0.06	-0.05	-0.07	-0.12	-0.11	-0.12	0.04	0.06	0.04	0.06
2-3	0.01	0.01	0.17	0.16	0.03	0.05	0.02	0.03	0.11	0.13	0.07	0.07	0.01	-0.01	0.01	-0.01
4-9	0.02	0.04	-0.09	-0.15	0.02	0.03	0.04	0.02	-0.02	-0.01	0.07	0.08	-0.02	-0.01	-0.02	-0.01
10-20	-0.03	0.00	0.00	0.00	0.00	0.03	0.03	0.04	-0.07	-0.03	0.00	0.00	-0.09	-0.10	-0.09	-0.10
21+	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<b>Number of offences 24 months prior</b>																
0-1	-0.03	-0.04	-0.10	-0.05	-0.04	-0.08	-0.06	-0.06	-0.07	-0.13	-0.11	-0.12	0.05	0.05	0.05	0.05
2-3	0.02	0.03	0.14	0.13	0.02	0.05	0.00	0.01	0.09	0.13	0.05	0.07	-0.01	-0.02	-0.01	-0.02
4-9	0.01	0.01	-0.06	-0.13	0.02	0.03	0.07	0.05	0.00	0.01	0.09	0.09	-0.02	-0.01	-0.02	-0.01
10-20	0.00	0.00	0.07	0.08	0.02	0.02	0.02	0.02	-0.08	-0.05	0.00	0.00	-0.02	-0.03	-0.02	-0.03
21+	0.00	0.00	0.00	0.00	0.03	0.07	0.00	0.00	0.08	0.09	0.00	0.00	0.00	0.00	0.00	0.00
<b>Prior offence severity (seriousness score)</b>																
Most severe offence 12m	0.15	0.17	0.10	0.06	0.12	0.16	0.15	0.17	0.09	0.07	0.15	0.16	0.09	0.03	0.09	0.03
Most severe offence 24m	0.23	0.26	0.11	0.08	0.20	0.24	0.22	0.24	0.13	0.12	0.15	0.16	0.15	0.15	0.15	0.15
Total offence severity 12m	0.11	0.18	0.07	0.00	0.10	0.16	0.18	0.18	0.05	0.04	0.14	0.18	0.03	-0.01	0.03	-0.01
Total offence severity 24m	0.20	0.24	0.12	0.07	0.19	0.24	<b>0.25</b>	<b>0.25</b>	0.07	0.08	0.16	0.18	0.09	0.09	0.09	0.09
Maximum prior severity	<b>0.25</b>	<b>0.27</b>	0.14	0.11	0.22	<b>0.26</b>	<b>0.25</b>	<b>0.26</b>	0.14	0.14	0.15	0.17	0.20	0.20	0.20	0.20
<b>Number of prior incidents/tasks</b>																
0	0.01	0.04	0.07	0.07	0.01	0.01	0.01	0.01	0.02	0.02	0.00	0.02	0.00	0.00	0.00	0.00
1	0.09	0.10	0.12	0.08	0.11	0.09	0.04	0.04	0.09	0.11	0.05	0.06	0.09	0.17	0.09	0.17
2-3	-0.01	0.01	-0.07	-0.08	-0.05	-0.02	-0.03	0.00	-0.01	-0.02	-0.05	-0.06	0.02	0.04	0.02	0.04
4-8	-0.04	-0.05	-0.03	0.00	-0.05	-0.03	-0.02	-0.03	-0.03	-0.05	-0.02	-0.04	-0.08	-0.11	-0.08	-0.11
9-20	-0.07	-0.10	-0.05	-0.03	-0.02	-0.06	0.01	-0.03	-0.05	-0.06	0.05	0.06	0.00	-0.05	0.00	-0.05
21+	0.03	0.05	-0.05	-0.06	0.02	0.03	0.03	0.04	0.00	0.00	0.00	0.00	-0.03	-0.09	-0.03	-0.09

**Age at first recorded offence**

<10	0	0	0	0	0	0	0	0	0	0.07	0.14	0.00	0.07	0.10	0.13	0.10	0.13
10-12	0	0	0	0	0	0	0	0	0	-0.01	0.04	0.04	0.09	-0.01	-0.07	-0.01	-0.07
13	0	0	0	0	0	0	0	0	0	0.10	0.04	0.16	0.09	0.01	0.02	0.01	0.02
14	0	0	0	0	0	0	0	0	0	0.06	0.04	-0.01	-0.01	0.02	0.03	0.02	0.03
15-17	0	0	0	0	0	0	0	0	0	-0.16	-0.18	-0.17	-0.19	-0.08	-0.03	-0.08	-0.03

**Most serious prior Police Intervention**

Other	0.02	0.05	0.00	0.00	0.01	0.03	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Warning	0.00	0.02	0.05	0.00	0.01	0.00	-	-	0.00	0.00	0.02	0.02	0.00	0.00	0.00	0.00	0.00
AA	-0.06	-0.03	0.04	0.00	-0.07	-0.06	-	-	0.08	0.06	-0.01	-0.01	0.02	0.07	0.02	0.07	
FGC	0.09	0.08	0.03	0.04	0.07	0.10	0.06	0.08	0.02	0.00	0.03	0.04	0.09	0.07	0.09	0.07	
Prosecution	0.00	-0.06	-0.13	-0.04	0.03	-0.01	0.05	-0.02	-0.14	-0.09	-0.03	-0.04	-0.11	-0.14	-0.11	-0.14	

**Police Intervention for reference proceeding**

Other	0.05	0.08	0.00	0.00	0.02	0.02	-	-	0.00	0.02	0.03	0.03	0.00	0.03	0.00	0.03
Warning	0.02	0.03	0.12	0.09	0.07	0.03	-	-	0.05	0.04	0.08	0.11	0.02	0.06	0.02	0.06
AA	-0.03	0.00	-0.02	-0.05	-0.06	-0.07	0	0	0.09	0.08	-0.07	-0.08	0.09	0.09	0.09	0.09
FGC	0.06	0.06	0.00	0.00	0.01	0.06	-	-	0.00	-0.02	0.00	0.00	0.00	0.02	0.00	0.02
Prosecution	-0.12	-0.20	-0.10	0.00	-	-	-	-	-0.24	-0.21	-0.04	-0.04	-0.22	<b>-0.28</b>	-0.22	<b>-0.28</b>

**Most serious prior Care and Protection (C&P) involvement**

None	-0.06	0.00	-0.03	-0.04	-0.09	-0.07	-0.11	-0.09	-0.05	-0.01	-0.07	-0.04	-0.10	-0.06	-0.10	-0.06
Intake	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Investigation	0.04	0.06	0.03	-0.03	0.01	0.02	0.04	0.01	0.05	0.03	0.06	0.00	0.09	0.11	0.09	0.11
FGC	0.03	0.00	0.00	0.06	0.07	0.07	0.07	0.07	-0.02	-0.07	0.02	0.05	0.06	0.03	0.06	0.03
Placement	0.02	-0.06	0.03	0.04	0.07	0.00	0.08	0.06	0.05	0.05	0.04	0.04	0.00	-0.05	0.00	-0.05

**Demographic/other**

Age at reference date	-0.10	-0.10	-0.07	-0.05	-0.08	-0.08	-0.07	-0.07	-0.23	-0.32	-0.17	-0.19	-0.18	-0.15	-0.18	-0.15
Female	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Māori	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Pacific Peoples	0.15	0.21	0.09	0.09	0.11	0.10	0.09	0.13	0.12	0.13	0.24	0.22	0.02	0.08	0.02	0.08

**Source:** MoJ (2024), Police (2023), Oranga Tamariki (2024). **Notes:** The table presents the standardised difference between the treatment and matched control group by specification and reoffending window for each matching variable. Unbalanced items are in **bold**.

Table B6: Balance table by number of nearest neighbours

	Months from reference date (reoffending window)					
Minimum number of neighbours	2		3		4	
Matching variables/Months	12	24	12	24	12	24
<b>Number of offences 12 months prior</b>						
0-1	-0.01	-0.02	-0.01	-0.04	-0.02	-0.04
2-3	0.04	0.03	0.06	0.05	0.06	0.06
4-9	-0.02	0.00	-0.04	0.00	-0.03	0.00
10-20	-0.06	-0.04	-0.07	-0.05	-0.08	-0.08
21+	0.00	0.00	0.00	0.00	0.00	0.00
<b>Number of offences 24 months prior</b>						
0-1	-0.01	-0.03	-0.02	-0.05	-0.03	-0.05
2-3	0.02	0.03	0.04	0.04	0.03	0.03
4-9	-0.01	0.01	-0.01	0.02	0.00	0.04
10-20	-0.02	-0.03	-0.04	-0.03	-0.04	-0.06
21+	0.03	0.07	0.03	0.04	0.02	0.05
<b>Prior offence severity (seriousness score)</b>						
Most severe offence 12m	0.13	0.14	0.15	0.17	0.14	0.16
Most severe offence 24m	0.20	0.22	0.23	0.25	0.22	0.24
Total offence severity 12m	0.07	0.11	0.07	0.12	0.06	0.09
Total offence severity 24m	0.15	0.19	0.16	0.20	0.15	0.17
Maximum prior severity	0.23	0.24	<b>0.25</b>	<b>0.27</b>	0.24	<b>0.26</b>
<b>Number of prior incidents/tasks</b>						
0	0.01	0.01	0.01	0.00	0.00	0.01
1	0.12	0.11	0.13	0.11	0.12	0.12
2-3	-0.04	-0.03	-0.05	-0.02	-0.02	-0.02
4-8	-0.05	-0.05	-0.06	-0.07	-0.08	-0.08
9-20	-0.05	-0.05	-0.04	-0.04	-0.03	-0.02
21+	0.07	0.06	0.07	0.05	0.06	0.01
<b>Age at first recorded offence</b>						
<10	0	0	0	0	0	0
10-12	0	0	0	0	0	0
13	0	0	0	0	0	0
14	0	0	0	0	0	0
15-17	0	0	0	0	0	0
<b>Most serious prior Police Intervention</b>						
Other	0.01	0.03	0.01	0.05	0.02	0.07
Warning	0.01	0.00	0.02	0.00	0.02	0.01
AA	-0.02	0.01	-0.02	-0.02	-0.02	-0.03
FGC	0.07	0.05	0.07	0.07	0.07	0.08
Prosecution	-0.05	-0.08	-0.05	-0.07	-0.06	-0.08
<b>Police Intervention for reference proceeding</b>						
Other	0.03	0.05	0.04	0.07	0.05	0.08
Warning	0.05	0.04	0.05	0.05	0.06	0.06
AA	0.00	0.02	-0.01	-0.02	-0.01	-0.03
FGC	0.02	0.02	0.02	0.03	0.02	0.04
Prosecution	-0.14	-0.19	-0.15	-0.16	-0.16	-0.17

**Most serious prior Care and Protection (C&P) involvement**

None	-0.05	-0.03	-0.08	-0.07	-0.09	-0.08
Intake	0.00	0.02	0.00	0.03	0.01	0.05
Investigation	0.03	0.04	0.05	0.05	0.07	0.07
FGC	0.03	0.02	0.04	0.04	0.03	0.03
Placement	0.03	-0.02	0.03	0.00	0.02	0.00

**Demographic/other**

Age at reference date	-0.07	-0.07	-0.06	-0.05	-0.06	-0.05
Female	0	0	0	0	0	0
Māori	0	0	0	0	0	0
Pacific Peoples	0.15	0.17	0.15	0.18	0.16	0.20

**Source:** MoJ (2024), Police (2023), Oranga Tamariki (2024). **Notes:** The table present the standardised difference between the treatment and matched control group by the number of (minimum) nearest neighbours for each reoffending window for each matching variable. Unbalanced items are in **bold**.

# Appendix C

This appendix summarises the findings from a supplementary analysis to the evaluation’s main findings.

First, the main findings were based on matching each participant with at least (and usually) one neighbour, which recorded the greatest similarity in terms of observables. Table C1 examines the impact of the pilot while relaxing this specification by increasing the minimum number of neighbours each participant must be matched with. Overall, the table shows that requiring a minimum of 2, 3, and 4 matches per participant shows statistically significant impacts on the likelihood to not reoffend (after 24 months) and to reduce reoffending frequency (after both 12 and 24 months), though these seem to reduce with each additional ‘neighbour’ (but not statistically different than when matching with a single neighbour). In addition, reductions in total offence severity were identified when matching participants with two (in either reoffending window) and three (in the 12-month window) neighbours.<sup>57</sup>

*Table C1: Social bond pilot impact estimates by outcome measure and reoffending period, by (minimum) number of neighbours*

Reoffending window	Frequency of reoffending		Severity of reoffending	
	None	Total	Max	Total
<b>Two neighbours</b>				
12 months (n=471)	0.023	0.100***	0.027	0.054**
24 months (n=392)	0.111***	0.112***	0.026	0.062**
<b>Three neighbours</b>				
12 months (n=471)	0.032	0.093***	0.026	0.049**
24 months (n=392)	0.091***	0.092***	0.012	0.046
<b>Four neighbours</b>				
12 months (n=471)	0.025	0.087***	0.023	0.046*
24 months (n=392)	0.083***	0.082***	0.012	0.043

**Source:** MoJ (2024), Police (2023), Oranga Tamariki (2024). **Notes:** Differences measured as the Average Treatment Effect on the Treated. Stars indicate the level of statistical significance (\* = 10%; \*\* = 5%; \*\*\* = 1%).

Next, as discussed in the limitations section, one inconsistency in the study design is the difference in reference dates for participants and the matched control group. For participants, this was their enrolment date, while for non-participants it was their proceeding date. As discussed, we used these dates as they are the best approximation of treatment start date. Table C2 uses the reference proceeding of participants as their reference date. While this possibly introduces biases to the estimation, it results in closer alignment between the groups in terms of time from apprehension. The table shows similar impacts in terms of reduced reoffending frequency (in both periods), while no statistically significant impacts over the likelihood to not offend and to reduce reoffending severity were detected.

<sup>57</sup> In terms of balance, Table B6 finds unbalance between the groups, as participants were more likely to record a greater maximum prior offence severity in the 3-neighbour sample in both periods, and in the 4-neighbour sample in the 24-months reoffending window.

*Table C2: Social bond pilot impact estimates by outcome measure and reoffending period, using reference-proceeding as reference date for participants*

Reoffending window	Frequency of reoffending		Severity of reoffending	
	None	Total	Max	Total
12 months (n=471)	0.011	0.079***	0.008	0.050*
24 months (n=392)	0.049	0.094***	0.014	0.04

**Source:** MoJ (2024), Police (2023), Oranga Tamariki (2024). **Notes:** Differences measured as the Average Treatment Effect on the Treated. Stars indicate the level of statistical significance (\* = 10%; \*\* = 5%; \*\*\* = 1%).

Next, as previously mentioned, there were relatively large differences (albeit still balanced) in the offending severity and proceeding outcomes histories between the participants and their matched control group. These differences will be taken into consideration in the following specifications.

First, Table C3 re-estimates the impacts of the bond when excluding the roughly 1% of participants and 2% of the matched control group who were prosecuted in their reference proceeding. Similarly, Table C4 re-estimates the impacts by focusing solely on the over 80% of rangatahi (from either group) who recorded an AA diversion in their reference proceeding. Overall, the results are not qualitatively different than those found for the entire sample. Table C3 indicates an 8pp increase in the share of participants who did not reoffend (in the 12-months reoffending window), while combined, the tables suggest a 7-11pp (12-21%) increase in the likelihood of recording reduced reoffending frequency (varies by table and reoffending window).<sup>58</sup>

*Table C3: Social bond pilot impact estimates by outcome measure and reoffending period, excluding rangatahi who were prosecuted at reference proceeding*

Reoffending window	Frequency of reoffending		Severity of reoffending	
	None	Total	Max	Total
12 months (n=466)	0.013	0.082**	0.031	0.060**
24 months (n=387)	0.083**	0.111***	0.011	0.053

**Source:** MoJ (2024), Police (2023), Oranga Tamariki (2024). **Notes:** Differences measured as the Average Treatment Effect on the Treated. Stars indicate the level of statistical significance (\* = 10%; \*\* = 5%; \*\*\* = 1%).

*Table C4: Social bond pilot impact estimates by outcome measure and reoffending period, all rangatahi with AA outcome at reference proceeding*

Reoffending window	Frequency of reoffending		Severity of reoffending	
	None	Total	Max	Total
12 months (n=397)	-0.004	0.074**	0.007	0.062*
24 months (n=328)	0.056	0.092**	-0.005	0.043

**Source:** MoJ (2024), Police (2023), Oranga Tamariki (2024). **Notes:** Differences measured as the Average Treatment Effect on the Treated. Stars indicate the level of statistical significance (\* = 10%; \*\* = 5%; \*\*\* = 1%).

Next, the balance table also indicated that participants recorded greater offence severity histories (i.e., regardless of the measure or period) compared with their matched control group. For example, the most severe offence across all participants was on average 274, or about one third greater than the mean across the matched

<sup>58</sup> Again, Table B6 indicates imbalance in the sense that participants (relatively to their matched control groups) record greater prior offence severity for some measures.

control group. To explore the potential impact of this difference, Table C5 splits the study population based on whether the maximum prior offence severity (by reference date) was less/equal to, or more than the median participant value (score of 179.79). In both panels, the matching results are very similar, indicating that participants were about 11pp (19-22%) more likely to reduce their reoffending frequency, with no significant changes found in terms of reduced severity.

*Table C5: Social bond pilot impact estimates by outcome measure, reoffending period, and maximum lifetime offence severity*

Reoffending window	Frequency of reoffending		Severity of reoffending	
	None	Total	Max	Total
<b>Less than or equal to median (&lt;=197.79)</b>				
12 months (n=266)	-0.000	0.061	0.026	0.042
24 months (n=221)	0.034	0.109**	0.067	0.068
<b>More than median (&gt;197.79)</b>				
12 months (n=205)	0.074	0.077	0.057	0.069
24 months (n=171)	0.090*	0.106**	0.024	0.039

**Source:** MoJ (2024), Police (2023), Oranga Tamariki (2024). **Notes:** Differences measured as the Average Treatment Effect on the Treated. Stars indicate the level of statistical significance (\* = 10%; \*\* = 5%; \*\*\* = 1%).

Finally, while not found to be a significant factor on reoffending outcomes (Table A3), a far greater share of participants was processed in the Counties Manukau Police District. Table C6 estimates the impacts of the bond by restricting the study population to only include rangatahi who were processed in Counties Manukau for the offending related to their reference proceeding. The table does not identify any significant impacts in terms of severity, or the likelihood of not reoffending, while the finding that reoffending frequency fell in the 24-month reoffending period continues to be significant for this sub-sample as well. Here, the pilot is estimated to increase the share of participants that reduced their reoffending frequency (during the 24-month reoffending window) by 9pp, or by nearly 17%. In terms of balance, while no items were deemed unbalanced, the share of matched non-participants who were prosecuted at the reference proceeding was far larger.

*Table C6: Social bond pilot impact estimates by outcome measure and reoffending period, all rangatahi processed in Counties Manukau Police District at reference proceeding*

Reoffending window	Frequency of reoffending		Severity of reoffending	
	None	Total	Max	Total
12 months (n=323)	-0.046	0.031	-0.057	-0.007
24 months (n=264)	0.049	0.093**	-0.015	0.013

**Source:** MoJ (2024), Police (2023), Oranga Tamariki (2024). **Notes:** Differences measured as the Average Treatment Effect on the Treated. Stars indicate the level of statistical significance (\* = 10%; \*\* = 5%; \*\*\* = 1%).

So far, the different specifications examined the *unconditional* impact of the pilot on reoffending outcomes. For example, the reduction in reoffending frequency estimated so far is comprised of both changes in the incidence rate (i.e., do people reoffend at all) and frequency (does the number of offences committed reduce).

Conceptually, the reductions in offending frequency estimated so far could be driven by fewer participants reoffending, and/or fewer offences committed amongst those who reoffended. While quantifying the magnitude of each of these two ‘channels’ is



beyond the scope of this evaluation, Table C7 presents some findings from an initial exploration, showing the estimates for the sub-samples of rangatahi who reoffended in the 12-month and 24-month periods. In the 12 months following enrolment, the table suggests that participants who offended were 11pp (53%), 9pp (32%), and 13pp (46%) more likely to reduce their reoffending frequency, maximum severity, and total severity, respectively. In the 24-month period, the only significant impact was a 10pp (47%) increase in the likelihood of to reduce offending frequency amongst those who offended. However, the balance table for the 24-month reoffending window (Table B5), indicates unbalance in terms of a lower share of participants who recorded a prosecution at the reference proceeding, which in turn, may upward bias the estimated impact (i.e., overstating impacts).

*Table C7: Social bond pilot impact estimates by outcome measure and reoffending period for rangatahi who offended within 12 and 24 months*

Reoffending window	Frequency of reoffending	Severity of reoffending	
	Total	Max	Total
12 months (n=205)	0.106***	0.090**	0.127***
24 months (n=171)	0.102**	-0.071	0.002

**Source:** MoJ (2024), Police (2023), Oranga Tamariki (2024). **Notes:** Differences measured as the Average Treatment Effect on the Treated. Stars indicate the level of statistical significance (\* = 10%; \*\* = 5%; \*\*\* = 1%).

Next, the impact of the pilot is examined by using alternative outcome measures. The table below presents 12- and 24-month outcomes in terms of the mean number of offences, total offence severity (or seriousness score), maximum offence severity, and average offence severity.<sup>59</sup> Table C8 suggests that on average, participants committed nearly two (1.976) fewer offences. No statistically significant changes were estimated for participants in terms of the severity measures.

*Table C8: Social bond pilot impact estimates by outcome measure and reoffending period, alternative outcome measures*

Reoffending window	Frequency of reoffending	Severity of reoffending		
	Offences	Total	Max	Mean
12 months (n=471)	-0.380	-1.250	-12.656	35.918
24 months (n=392)	-1.976***	-205.799*	2.978	-2.420

**Source:** MoJ (2024), Police (2023), Oranga Tamariki (2024). **Notes:** Differences measured as the Average Treatment Effect on the Treated. Stars indicate the level of statistical significance (\* = 10%; \*\* = 5%; \*\*\* = 1%).

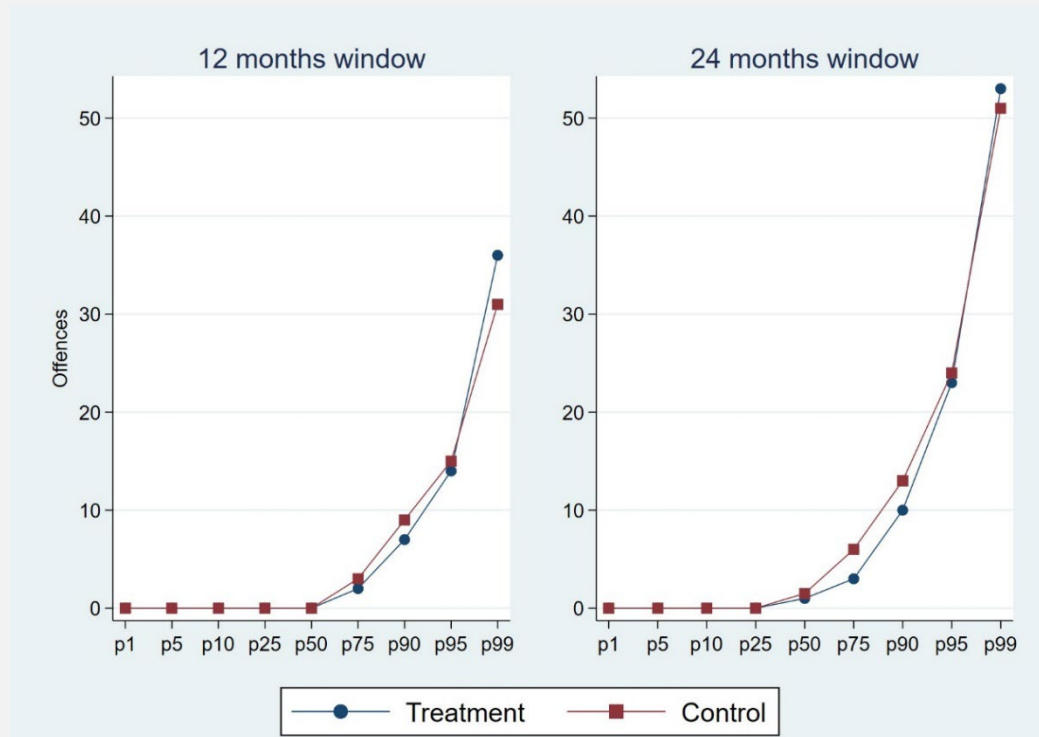
To better understand the underlying distribution comprising the estimated means in Table C8, Figure C1-Figure C3 present the distributions of reoffending, maximum offence severity, and total offence severity across rangatahi in the treatment and matched control groups. Figure C1 suggests that the lower reduction in the number of mean offences for participants is more likely to be driven by reductions in offence counts at the upper end of the offending distribution, with participants recording fewer offences in most points from the 75<sup>th</sup> percentile.

The patterns relating to severity are not as clear, with many instances at the upper end of the distribution where participants record greater values (e.g., 99<sup>th</sup> percentiles in offence severity figures, and 12-month total severity).

<sup>59</sup> If rangatahi did not reoffend, the severity measures are set to zero.

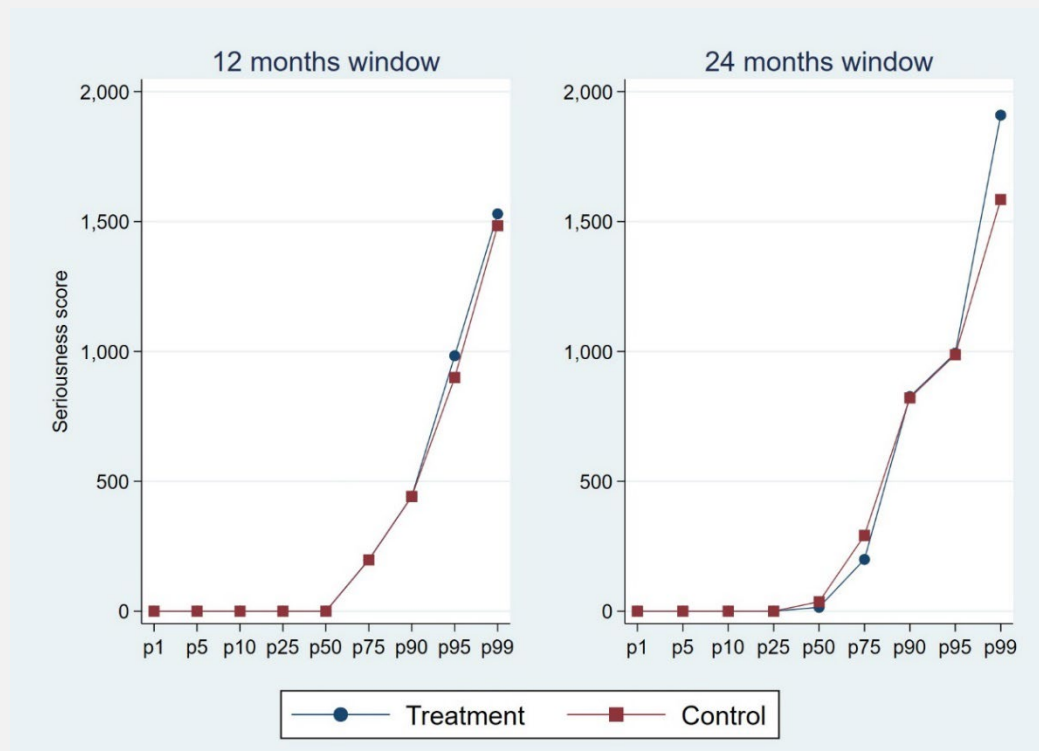
These findings, once again highlight the risk of using cohort level outcome measures (e.g., mean change) when evaluating the performance of the pilot, since the relatively small number of participants mean that such findings are highly sensitive to outliers.

Figure C1: Total offending distribution by treatment status and reoffending window



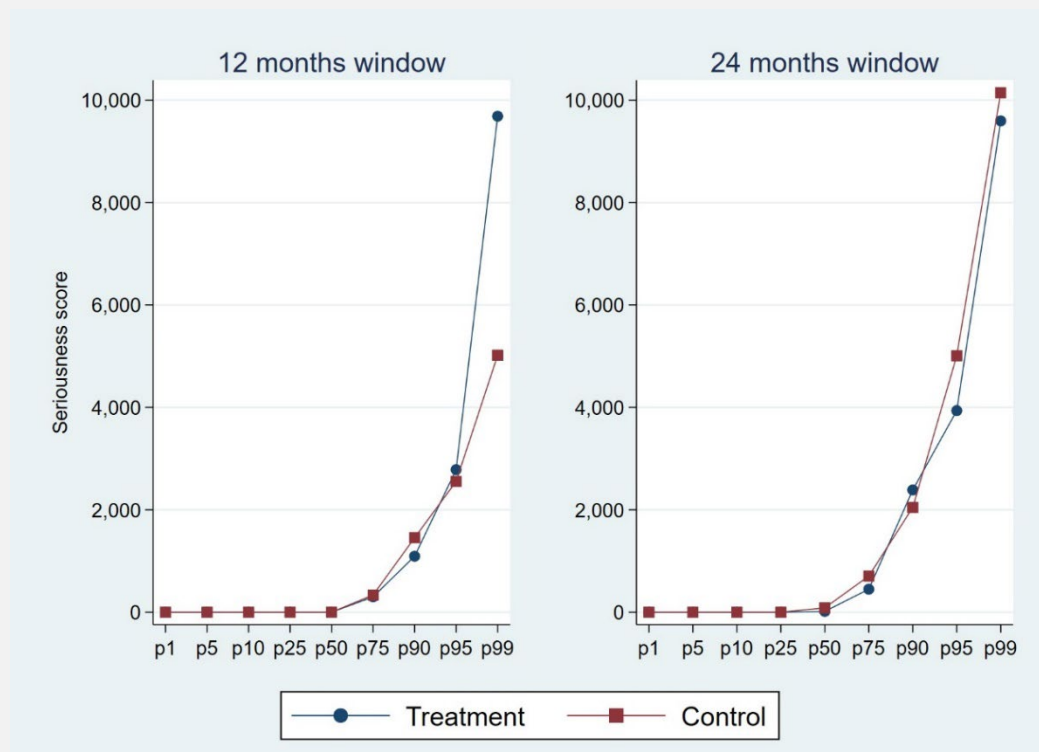
Source: MoJ (2024), Police (2023), Oranga Tamariki (2024). Note: Control group distribution is calculated without applying weights.

Figure C2: Maximum offence severity distribution by treatment status and reoffending window



Source: MoJ (2024), Police (2023), Oranga Tamariki (2024). Note: Control group distribution is calculated without applying weights.

Figure C3: Total offence severity distribution by treatment status and reoffending window



Source: MoJ (2024), Police (2023), Oranga Tamariki (2024). Note: Control group distribution is calculated without applying weights.

Finally, Table C9 re-estimates the impacts of the SB pilot applying a Propensity Score Matching (PSM) approach, and Table C10 while applying Inverse Probability Weighted Regression Adjustment (IPWRA). Using PSM, the results are similar to those using NNM, finding a 10pp increase in the likelihood of participants not reoffending, and reduced frequency (with no significant impact on severity). Using IPRWA, the pilot was found to increase the likelihood of participants not reoffending (7pp) after 24 months and increase the likelihood to record reduced frequency and severity in either period.

*Table C9: Propensity Score Match estimates, all participants*

Reoffending window	Frequency of reoffending		Severity of reoffending	
	None	Total	Max	Total
12 months (n=466)	-0.022	0.029	0.015	0.030
24 months (n=387)	0.105***	0.102***	0.055	0.048

**Source:** MoJ (2024), Police (2023), Oranga Tamariki (2024). **Notes:** Differences measured as the Average Treatment Effect on the Treated. Stars indicate the level of statistical significance (\* = 10%; \*\* = 5%; \*\*\* = 1%).

*Table C10: Inverse Probability Weighted Regression Adjustment estimates, all participants*

Reoffending window	Frequency of reoffending		Severity of reoffending	
	None	Total	Max	Total
12 months (n=466)	0.023	0.080***	0.029	0.053***
24 months (n=387)	0.072***	0.079***	0.043*	0.060**

**Source:** MoJ (2024), Police (2023), Oranga Tamariki (2024). **Notes:** Differences measured as the Average Treatment Effect on the Treated. Stars indicate the level of statistical significance (\* = 10%; \*\* = 5%; \*\*\* = 1%).