Tūtohi Wild Bamboo

Tamariki at the heart

Stories, statistics, and solutions from a quantitative evaluation of the Kōkihi ngā Rito programme

Creative Commons licence

Liability

Recommended citation

Tūtohi (2025). Tamariki at the heart: Stories, statistics, and solutions from a quantitative evaluation of the Kōkihi ngā Rito programme. Aotearoa New Zealand.

ISBN 978-0-473-74354-3 (online)

Authors

Jillian Pennington—Tūtohi, Wild Bamboo
Sandra McCallum—Tūtohi, Wild Bamboo
Juliet Shaw—Tūtohi, Wild Bamboo
Design and illustration by Stacey Tangaere (Ngāti Porou, Ngāpuhi)—Tūtohi, Wild Bamboo

Acknowledgements

We extend our deepest gratitude to the Kaiārahi Tamariki, the refuges, the whānau, and most importantly, the tamariki themselves, whose experiences are reflected in the data that informed this report. Although we did not speak with tamariki directly, we heard their voices echo through the operational data that was collected and ultimately shared for evaluation. To the tamariki and whānau whose lives are represented in this work, we are humbled by the opportunity to reflect your stories through this evaluation for the benefit of others. We hope that these analyses honour and amplify your voices in a way that does justice to your lived experiences and to the mahi of those who support you. To the Kaiārahi Tamariki and refuges, thank you for your unwavering dedication to supporting tamariki and whānau on their journeys toward safety and healing.

We also wish to acknowledge the gracious and insightful reviewers from MSD, Te Puna Aonui, NCIWR, Oranga Tamariki, and Wise Group who generously contributed their time, expertise, and insights to the development of this report. Your thoughtful feedback and constructive suggestions were invaluable in ensuring the clarity, accuracy, and relevance of our findings and we are deeply grateful for your support and collaboration in this mahi.

Finally, we extend our deep appreciation to the authors of the qualitative evaluation of the Kōkihi ngā Rito programme. Throughout this report, you will find the authentic voices of tamariki, mothers, Kaiārahi Tamariki, and NCIWR Refuge managers—all drawn from that foundational work. These powerful testimonies have guided and inspired our analysis, and the lived experiences of those at the heart of this programme remain central to our quantitative findings.

Contents

Executive Summary

Executive summary

A snapshot of insights and impact

New Zealand has the highest rate of family violence in the OECD, a dark shadow cast over individuals, whānau, and our communities. Its effects are far-reaching and devastating, impacting not just those directly harmed but also the wider fabric of our society. Every year, tens of thousands of wāhine seek safety and support through the National Collective of Independent Women's Refuges (NCIWR), often accompanied by tamariki whose lives have also been disrupted. For these children, family violence is not a reflection of adult struggles—it is a direct assault on their safety, wellbeing, and sense of self.

NCIWR has long been a lifeline for both wāhine and their tamariki, providing critical services and support in more than 40 independent refuges dotted across the country. Recognising the unique and often invisible scars carried by the child victims of family violence, NCIWR has continually sought to innovate care models to meet their needs.

Developed with the guidance and input of tamariki, the Kōkihi ngā Rito programme is a novel, child-centred model of advocacy that respects children as individuals with their own voices, needs, and aspirations. Delivered by specialist Kaiārahi Tamariki, Kōkihi ngā Rito helps tamariki to process their experiences, rebuild trust, and regain confidence. By offering sustained, relational care, the programme provides an alternative to traditional models that prioritise maternal experiences.

The programme also serves as a blueprint for change, setting a new standard for advocacy within NCIWR. Through its tailored supports—emotional, practical, and relational—it aims to empower tamariki on their individual journeys of healing and growth. It has also illuminated new pathways for systemic transformation.

This report provides a quantitative evaluation of the Kōkihi ngā Rito programme, delivered since 2021 in six NCIWR refuges nationally. We aimed to balance rigorous analysis with respectful understanding of the lived realities faced by tamariki and their mothers, measuring the impact of Kōkihi ngā Rito, its ripple effects within a refuge, and the broader implications it may have for family violence services.

The interweaving of information, analysis, and the voices of tamariki highlights the transformative potential of Kōkihi ngā Rito. It exposes the urgent need for expanded, sustained investment in services that centre our youngest survivors—offering them not just safety, but the opportunity to heal, thrive, and reclaim their futures.

70%

of risk assessments report that
a child has been harmed
by family violence

This translates to more than

25,000

tamariki annually

This is the reality for these tamariki:

1 in 2 have been threatened

1 in 3 have been physically harmed

1 in 3 have been kept from their mother

1 in 2 have witnessed their mother being physically abused

1 in 2 have witnessed their mother being verbally abused

1 in 4 have witnessed their mother being sexually abused

Kōkihi ngā Rito

A blueprint for change

Kōkihi ngā Rito transforms support for tamariki affected by family violence by providing dedicated, specialist child advocates (Kaiārahi Tamariki). Unlike many mainstream models where children's needs are secondary to their mothers', this programme centres tamariki voices and relationships, shifting from crisis-focused, short-term interventions to sustained, comprehensive support.

Kōkihi ngā Rito delivers significantly better outcomes in key wellbeing domains and elevates care standards across entire refuges, demonstrating how child-centred design can revolutionise support for our most vulnerable tamariki.

75%
of tamariki felt safer

69%
of tamariki felt more self-confident, with greater self-esteem

75%
of tamariki felt better able to deal with their feelings and behaviour

Tamariki flourish

Whānau thrive

Mothers heal

Services improve

When we invest in specialist support for the child victims of family violence, we don't just change statistics—we change lives

Next steps | Build insights into impact

Expand critical reach to vulnerable tamariki → **Close the support gap for those most at risk** → **Quantify economic and social value** → **Leverage evidence-based mechanisms of impact** → **Advocate for system-wide transformation**

Key findings

Five key findings in this report illuminate the measurable, life-changing impact of Kōkihi ngā Rito.

Tamariki wellbeing transformed

Tamariki supported by Kōkihi ngā Rito experienced remarkable improvements across all eight wellbeing domains measured by My Star™, with the most profound gains in *Being safe*, *Feelings and behaviour*, and *Confidence and self-esteem*. Most strikingly, their wellbeing improved significantly more than children supported by other NCIWR services, underscoring the programme's unique impact in enhancing wellbeing and creating safer futures for tamariki.

Mothers thrived alongside their tamariki

When tamariki are supported, their mothers also flourish. The wellbeing of mothers whose children were supported by NCIWR improved significantly in the specific domains of *Children* and *Work and learning,* reinforcing a deep interconnection between a child's safety and a mother's stability. These findings highlight the broad reach of child-centred support services—when we lift tamariki, we strengthen entire whānau.

More care meant better outcomes

Kōkihi ngā Rito did more than just provide support—it redefined service delivery. Tamariki in the programme spent twice as long in service, received 5-7 times more support hours (even corrected for referral duration), and had significantly longer and more detailed case notes compared to those in other services. The presence of a specialist Kaiārahi Tamariki also corresponded to a service delivery uplift across other services in a refuge, raising the standard of care for tamariki even beyond those enrolled in the programme.

Tamariki voices were heard, loud and clear

Case notes from Kōkihi ngā Rito captured tamariki perspectives with unprecedented depth, particularly around feelings, whānau, and violence. This reflects the programme's fidelity to an advocacy model that places tamariki at the heart—not as passive recipients of care, but as individuals whose experiences and voices shape the support they receive.

The scale of unmet need is stark but solvable

More than 70% of wāhine supported by NCIWR report that their tamariki have experienced or been exposed to family violence, which corresponds to a conservative estimate of over 25,000 tamariki each year. These children have been threatened with harm, physically abused, separated from their mother, forced to witness verbal, physical, and even sexual abuse—yet fewer than 1 in 5 are currently able to be supported by NCIWR, and less than 1% have access to Kōkihi ngā Rito. These children are not statistics—they are known, named, and they are in need.

Recommendations

Programmes like Kōkihi ngā Rito offer an opportunity to reset the future trajectories for our most vulnerable tamariki—a chance to rewrite stories of trauma into ones of resilience and hope. Intervening early with targeted, child-centred care could transform not just individual lives, but also the systems and communities that sustain them. Therefore, the following recommendations are proposed:

Expand critical reach to vulnerable tamariki

- **Deploy specialist Kaiārahi Tamariki** across additional refuges to create a safety net for more at-risk tamariki
- **Use risk assessment data to inform community engagement** by analysing service patterns and community needs to guide resource allocation and develop responsive outreach strategies that connect with tamariki who may benefit most from support
- **Extend vital support to rangatahi** by expanding programme eligibility beyond age 12, a critical service gap for adolescents

Close the support gap for those most at risk

- **Scale infrastructure and expertise** through targeted investment in specialised staff (maintaining the rigorous training standards of the programme) and enhanced service delivery systems to reach those tamariki who are experiencing harm without support
- **Strengthen strategic partnerships** with government agencies, funders, and community stakeholders to secure dedicated resources for child-focused family violence interventions

Quantify economic and social value

- **Conduct a comprehensive Social Return on Investment (SROI)** analysis that captures the full economic impact of Kōkihi ngā Rito
- **Centre tamariki perspectives in value assessment** through meaningful consultation to determine financial values for safety and wellbeing outcomes—an approach that will address the exclusion of children's voices from impact valuation frameworks

Leverage evidence-based mechanisms of impact

- **Investigate the observed connection** between tamariki support and maternal wellbeing improvements, particularly within the *Children* and *Work and Learning* domains
- **Identify the mechanism of transformative differences** in service delivery at a refuge when specialist Kaiārahi Tamariki are present, using mixed-methods research and kaimihi consultation
- **Develop replication strategies** that preserve the integrity of the core Kōkihi ngā Rito service model while extending successful impact mechanisms across the broader refuge network

Advocate for system-wide transformation

- **Develop specialised guidance** for professionals across family court, education, health, and social services to improve interactions with child victims of family violence
- **Champion children's right to representation** by ensuring every tamaiti has access to a trained, specialist child advocate, so that tamariki voices are meaningfully included in other decision-making processes that impact them, particularly in the family court system.

These recommendations are specifically linked to the Kōkihi ngā Rito service model. The positive outcomes documented in this evaluation reflect the specialist approach delivered by those extensively trained Kaiārahi Tamariki and we caution against extrapolating those results more broadly to other children's programmes. For any initiatives seeking to replicate these outcomes, we encourage fidelity to the comprehensive training, supervision, and support systems of Kōkihi ngā Rito.

Conclusions

The findings of this quantitative evaluation underscore the profound and measurable impact of the Kōkihi ngā Rito programme. Tamariki who participate in the programme experience significant improvements in wellbeing, surpassing gains seen in other services. But the impact extends even further—into the lives of their mothers, into the wider refuge environment, and into the very fabric of how care is delivered. These results are not merely luck—they reflect the intentional, evidence-based design of the programme and its deep commitment to child-centred, relational advocacy that puts tamariki at the heart of care.

The data tell a story of transformation. Kōkihi ngā Rito tamariki are seen, heard, and supported in ways that extend beyond traditional service models. They receive more time, more care, and more meaningful support than those in traditional service models. Their voices are captured with unprecedented depth in case notes, ensuring that their experiences shape the support they receive. The ripple effects of this model reach beyond the programme itself, elevating service delivery across entire refuges, lifting the standard of care even for tamariki outside the programme. This is systemic change in action.

Yet, these findings also expose a stark and urgent truth. Tens of thousands of tamariki are connected to wāhine supported by NCIWR each year, but fewer than 1% receive the depth of care offered by Kōkihi ngā Rito. Risk assessment data reveal a devastating reality: too many children are facing severe and enduring harm without access to the specialised support they need. This disparity is not just a gap—it is a crisis. It demands more than recognition; it demands action.

The evidence is clear and compelling: Kōkihi ngā Rito is not just delivering measurable outcomes, but meaningful change. Investing in tamariki is an investment in futures, in whānau, and in communities. Kōkihi ngā Rito makes a difference for tamariki today, but this targeted, early intervention for the child victims of family violence also builds a foundation for the safer, stronger communities of tomorrow. The challenge now is not in proving the model—it is expanding its reach and scaling the impact, ensuring that the rising tide of Kōkihi ngā Rito lifts every child it touches, including the many more who are still waiting.

If we are not going to advocate for tamariki, who will?

- Kaiārahi Tamariki

Background

Background
Why tamariki voices matter

Family violence in Aotearoa

New Zealand carries the grim distinction of having the highest rate of family violence anywhere in the OECD[1]. Behind the statistic are countless stories of harm—and most go untold. Seventy percent of family violence incidents occur in households with children[2]. One in four children in the Dunedin Study witnessed threats or abuse between parents while growing up[3] and another study found that 40% of children reported at least one violent act by a parent, most before the age of 11[4]. Inside a year, one in seven children witness physical violence at home, while even more become the unwilling audience to verbal abuse[4]. Police respond to a family violence call out every three minutes[5], and, heartbreakingly, a child dies due to family violence in Aotearoa approximately every five weeks[6].

The effects of exposure to family violence during childhood are profound and far-reaching, impairing cognitive ability[7,8] and school achievement[9,10,11,12], eroding emotional and social wellbeing[12,13,14,15], and even altering the physical development of a child's brain[16,17,18]. This harm is cumulative—a cruel arithmetic of adversity—with greater exposure leading to deeper wounds, yet these injuries are often hidden during a "silent period", a delay between maltreatment and the emergence of its effects[19]. The lack of immediate, observable effects often leads adults to mistakenly believe a child is unharmed or unaffected by exposure to family violence, when in reality, it is common to see effects manifest months or even years later[20,21,22,23].

Importantly, the harm is not limited to acts of physical violence. The sight of verbal abuse, the sound of threats—these too are weapons that wound. A child's distress is equally as sharp and damaging, whether violence is enacted or merely threatened[24,25]. Fear itself leaves scars, and family violence doesn't have to target children directly to harm them.

He hit across my head and locked me outside.

- Boy, 7 years old

I'm checking the locks every night.

- Boy, 9 years old

I don't want Mum to die.

- Girl, 10 years old

In Aotearoa, **a child dies** due to family violence **every five weeks**

A history of Women's Refuge services in Aotearoa

Against this backdrop, Women's Refuge has stood steadfast for over 50 years, a vital lifeline for those escaping violence. What began as grassroots activism in 1973 with the opening of Aotearoa's first Women's Refuge in Ōtautahi Christchurch has grown into a network of more than 50 refuges across the country. Forty-one of these refuges are affiliated with the National Collective of Independent Women's Refuges (NCIWR), an umbrella organisation established in 1981 to support and coordinate member refuges.

Today, these 41 refuges span the motu, from Kaitaia to Invercargill, working together to provide consistent, high-quality services while retaining their independence to respond to local needs[26]. These refuges share a unified database for recording client information, ensuring consistency and collaboration across the network while maintaining the autonomy of each refuge to address the varying needs of its local area. Each refuge is as unique as the community it serves, but together they form a web of safety and advocacy. Every year, the NCIWR refuges support over 25,000 women and children, providing essential services to help them rebuild their lives.

Their core purpose of liberating women, children and whānau from family violence is possible only when autonomy and agency are restored to those who have been controlled, silenced, or harmed. This model is underpinned by tools like the Empowerment Star™ [27], a framework that helps wāhine identify needs, set goals, and map progress in reclaiming their lives.

Alongside this, NCIWR places a significant emphasis on understanding and mitigating risk, with a perspective of risk that spans beyond threats to immediate physical safety to include the full range of (often unseen) risks to women's lives, futures, and wellbeing[28]. Comprehensive risk evaluations are conducted to assess the safety needs of wāhine and tamariki, informed by decades of research, data analysis, and frontline expertise. The organisation's policy work, grounded in these insights, shapes national conversations and legislative responses to family violence.

The work of NCIWR is multifaceted and strives to address not only the immediate need for physical safety but also the broader and more enduring work of emotional healing, social reconnection, and rebuilding self-worth for victims of family violence.

WOMEN'S REFUGE

Kōkihi ngā Rito
A new approach to child-led advocacy

In July 2021, NCIWR, in partnership with the Ministry of Social Development, launched the Kōkihi ngā Rito programme—a new and groundbreaking initiative designed to place tamariki at the centre of care[29]. Among the first of its kind, both here in Aotearoa and internationally, this approach to child-led advocacy was designed in response to the Kids in the Middle research[30], which put children in the role of expert advisors, asking them directly what they need from family violence services.

Where traditional mainstream responses to family violence have often focused on adult victims, Kōkihi ngā Rito acknowledges the distinct and significant harm experienced by children, not as collateral damage but as victims in their own right.

Kaiārahi Tamariki work within a framework that is both structured and fluid, informed by their extensive skills and knowledge. They bring not only technical expertise but also a deep understanding of relational practice, allowing them to connect meaningfully with tamariki and adapt their approach to each unique situation. The 'wavering door' principle ensures that support does not end until a child feels their needs have been met, reflecting the programme's deep respect for the agency and voice of tamariki. The My Star™[31] tool helps children articulate their feelings, set goals, and track their progress in ways that feel meaningful to them. But it is the relational nature of the work—the quiet patience, the gentle persistence, the attunement to unspoken needs—that forms the heart of the programme. Kaiārahi Tamariki earn the trust of tamariki, often over many months, creating the safe spaces necessary for healing to begin.

While Kōkihi ngā Rito focuses on the unique needs of tamariki, it also works to support and strengthen whānau. Caregivers are supported to understand the lingering impacts of violence, recognise the specific needs of their children, and develop strategies to foster safety and growth. This holistic approach reflects the interconnectedness of whānau, recognising that the wellbeing of tamariki is deeply tied to the environments in which they live and grow.

Counsellors ask so many questions, but [my Kaiārahi Tamariki] just listens and understands.

- Tamaiti

Methods

Methods

How quantitative methods helped us listen

The data used in this evaluation came from information routinely collected during the services that NCIWR provides, and was not gathered specifically for the purpose of this analysis. Because of this, we had to carefully prepare the data and make thoughtful decisions about how to analyse it. These decisions are important for understanding the findings, and we've included full details in the technical report and appendices for anyone wanting to dive deeper.

How we sourced the data and kept it safe

Data was collected by NCIWR staff within their Recordbase client management system, with informed consent from service users, and then securely transferred to a separate database for analysis. In line with New Zealand government's Cloud First policy, all data was stored within a secure Microsoft Azure environment that undergoes regular penetration testing as well as security audits by cybersecurity experts. All members of the analysis team undergo MOJ and MSD vetting as standard, and Azure AD single sign-on (SSO) with multi-factor authentication (MFA) was used to allow named users access to the data, with user roles governed by the principle of least privilege. All data was deidentified and required portions were then extracted for downstream analysis using Stata and DataTab, with visualisation in PowerBI. All code underwent independent peer review and all statistical analyses were independently repeated for verification.

Who we included and excluded

To compare the impact of different types of support, we grouped tamariki and wāhine into cohorts based on the type of support they (or their children) received at NCIWR. The groups, shown in the following diagram, were:

Tamariki cohorts

- Tamariki supported by Kōkihi ngā Rito
- Tamariki supported by other services at a refuge that also delivers Kōkihi ngā Rito
- Tamariki supported by refuges that do not deliver the Kōkihi ngā Rito programme

Wāhine cohorts

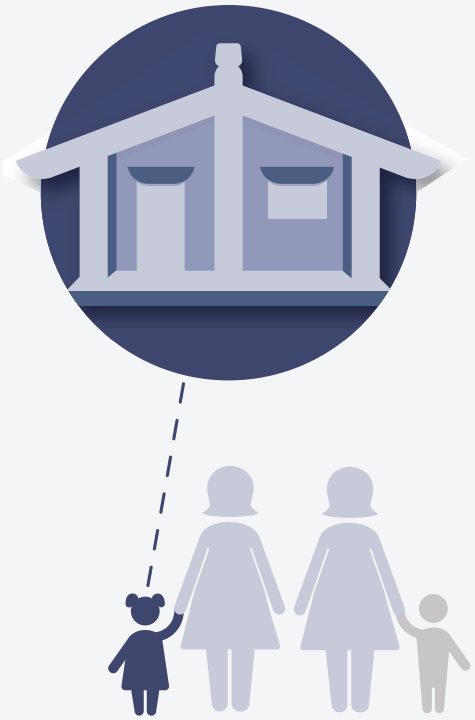
- Mothers whose children (all of them) were supported only by the Kōkihi ngā Rito programme
- Mothers whose children (all of them) received other types of NCIWR support (at any refuge)
- Mothers whose children (all of them) received no direct support from NCIWR

Cohort inclusion and exclusion criteria

This refuge delivers Kōkihi ngā Rito

Some children whose mothers are supported at this refuge will be enrolled in Kōkihi ngā Rito, while others may be supported by other community or local services at the same refuge.

This refuge does not deliver Kōkihi ngā Rito



Some children whose mothers are supported at this refuge may also be supported by other community or local services at this refuge.

Unfortunately, not all children can be supported at either refuge.

Tamariki cohorts

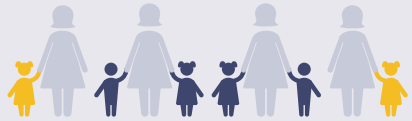
Tamariki supported by Kōkihi ngā Rito

Tamariki supported by other services at a refuge with Kōkihi ngā Rito

Tamariki supported by refuges that do not offer Kōkihi ngā Rito

Wāhine cohorts

Children supported only by Kōkihi ngā Rito



Children supported only by other services (at any refuge)

Children received no direct support from NCIWR

How do we measure wellbeing?

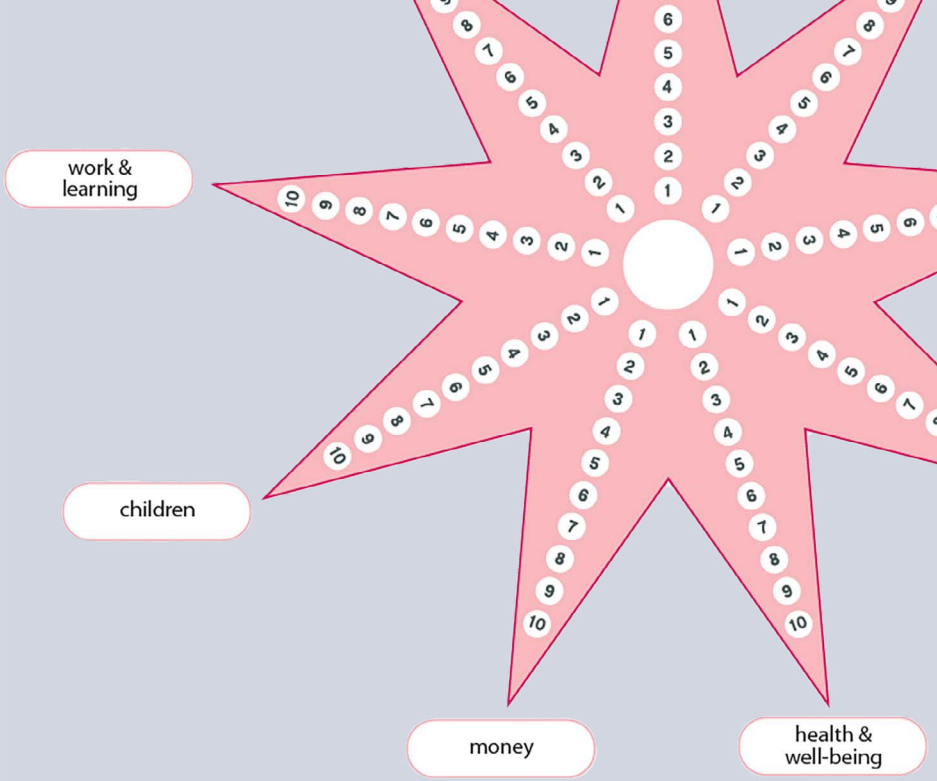
The Outcomes Star™, developed by Triangle Consulting, is a family of evidence-based tools designed to support and measure change[32]. Grounded in research, with robust psychometric validation, these collaborative tools serve not only as a measure of progress but as catalysts for achieving meaningful outcomes.

Each tool in the Outcomes Star™ family focuses on key domains of wellbeing that are tailored to specific populations. Clients score their progress within each domain, using scales based on a five- or ten-point model of change. Each scale outlines specific behaviours and attitudes at every stage, encouraging reflection and dialogue. Variants like the Family Star™, Teen Star™, and Youth Star™ address the unique needs of different groups, but all versions share the same purpose: helping individuals to reflect on and navigate their journey toward change.

The Outcomes Star™ tools align strongly with NCIWR principles by challenging the traditional power dynamics of evaluation. By centring service user voice and agency rather than prioritising external clinical judgements, these tools support a collaborative process that encourages service users to shape, guide, and own both their support needs and their individual journeys.

To ensure the consistent and correct use of these tools, every practitioner must complete specialised core training before receiving a licence to use them, and NCIWR National Office provides ongoing sector-specific support and supervision.

This quantitative evaluation incorporates data from two Stars used by NCIWR: the Empowerment Star™, and the My Star™. The datasets from these tools provide a structured, evidence-based framework to measure progress and understand the unique experiences of wāhine and tamariki navigating change and healing.



The **Empowerment Star™**[27] was designed for women who are victims of domestic violence. It includes nine domains of wellbeing, each scored from 1 (Not ready for help) to 10 (Independence and choice):

- Safety
- Accommodation
- Support networks
- Legal issues
- Health and wellbeing
- Money
- Children
- Work and learning
- Empowerment and self-esteem

The **My Star™**[31] was designed for children and young people in families identified as vulnerable, in settings such as children's homes or foster care, or facing issues such as bereavement or abuse. Recognising that a child's wellbeing may be profoundly impacted by factors outside their control, the My Star™ includes two scoring sections, each with four domains:

The ***Journey of Change*** is used for those areas where change depends mostly on the child learning new skills or doing things differently, scored from 1 (stuck) to 5 (resilient):

- Feelings and behaviour
- Friends
- Confidence and self-esteem
- Education and learning

The ***Severity Scale*** is used for those aspects of life where it is most important to know how difficult or otherwise things are for the child, where improvement depends on a change in the child's circumstances, scored from 1 (big concerns) to 5 (things are good):

- Physical health
- Where you live
- Being safe
- Relationships

The data we used

Wellbeing data

For wellbeing comparisons, we used data from the My Star™ tool, which helps measure and track a child's progress[30]. However, My Star™ is not routinely used outside of Kōkihi ngā Rito, so we could only compare 71 tamariki from Kōkihi ngā Rito with 43 tamariki from other services—a very small subset of all tamariki referrals. These 43 children likely represent cases where staff chose to use My Star™ for children with higher needs, meaning they may be more similar to Kōkihi ngā Rito participants than other children in NCIWR services. For wāhine wellbeing, we used data from the comparable tool, the Empowerment Star™ [27].

Service delivery data

For service delivery comparisons, we extracted the dates and details of individual service episodes (often called "referrals") as well as durations of all support activities and the contents of all case notes. As free text, case notes could not be reliably deidentified, so the analysis team developed and tested all case note code using dummy data before applying it to the real data set. Because we did not require My Star™ data, the cohort sizes were much larger—155 tamariki supported by Kōkihi ngā Rito, 2,973 supported by other services at those same refuges, and 12,277 supported at other refuges.

Demographic data

To test whether outcomes were consistent across demographics and to ensure cohorts were similar in composition, we extracted age, ethnicity (at StatsNZ Level 1), gender, whānau composition (number of siblings), deprivation index, and mother's risk assessment data.

How we handled the data

Because human journeys through family violence and support can be complex and nuanced, we had to make some adjustments to ensure a fair comparison between tamariki and wāhine across the different groups:

1. For tamariki who received support more than once, we included all their service episodes but treated each as independent, as long as there was a discharge that separated them. This affected 2 tamariki.

2. To remove an obvious potential confounder, we included only wāhine where all their tamariki received the same type of support experience. This excluded 246 wāhine out of 3,380.

3. When analysing My Star™ data, we created two comparison groups to address overlap where tamariki had been supported by both Kōkihi ngā Rito and other services. All analyses were repeated with both groups to confirm that results were robust and did not change according to the researchers' decisions here. Only 12 tamariki were in this overlap group.

4. For service delivery comparisons, we repeated all analyses with both the entire group of tamariki and just the subset whose mothers' risk assessments indicated the presence of all six child-related harms, ensuring that comparisons were not sensitive to varying risk levels for tamariki. For the three cohorts (Kōkihi ngā Rito, other services at a refuge delivering Kōkihi ngā Rito, and refuges without Kōkihi ngā Rito), this adjusted the sample sizes from 155 to 56; 2,973 to 357; and 12,277 to 705; respectively.

Why these decisions matter

Understanding the decisions made during this evaluation is key to interpreting the results with clarity and confidence. These choices help to illuminate the limitations inherent in observational data—such as potential biases in cohort selection—while also identifying opportunities to enhance future analyses. Expanding the routine use of tools like My Star™ or systematically tracking service delivery metrics across programmes could strengthen the ability to evaluate services and improve outcomes for tamariki and their whānau.

At the same time, it is important to acknowledge the robust data collection protocols already in place that made these analyses possible. The thoughtful documentation practices of Advocates and Kaiārahi Tamariki, alongside the consistent use of detailed case records, provided a strong foundation for this evaluation. These records not only allow for the examination of patterns and outcomes but also reflect the care and intentionality of the support provided.

By thoughtfully addressing these challenges, this evaluation leverages the strengths of NCIWR data practices to provide meaningful insights. Although limitations are present in any analysis, the layering of multiple statistical approaches and analyses provide a robust, compelling foundation for understanding the impact of Kōkihi ngā Rito and other NCIWR services. The following section explores the results of these analyses, revealing how the data highlights the transformative potential of child-centred advocacy and its broader implications for tamariki, whānau, and service delivery.

Results

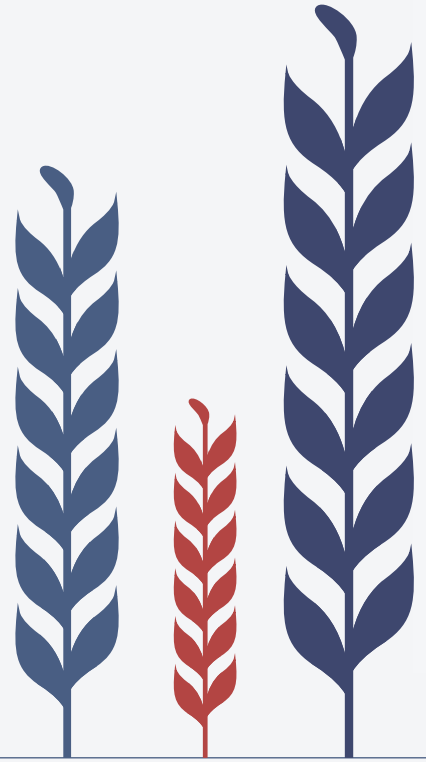
Results

The stories in the statistics

This section presents the key findings of the quantitative evaluation of Kōkihi ngā Rito, focusing on the programme's impact on tamariki wellbeing, service delivery, and broader sector outcomes. The analysis examines whether Kōkihi ngā Rito has achieved its intended outcomes and how its approach compares to other tamariki support services.

For accessibility, this report provides a high-level summary of the key results and what they mean. A more detailed statistical breakdown, including full tables of results and technical explanations of the analyses performed, is available in the accompanying technical report. Readers seeking in-depth methodological details, statistical assumptions, and effect size calculations are encouraged to refer to that document.

The findings in this section are structured around the core evaluation questions, providing an overview of what the data reveal about the programme's effectiveness. Each question is explored below, with results framed in practical terms to highlight their significance for tamariki, whānau, and service providers.



A journey of change

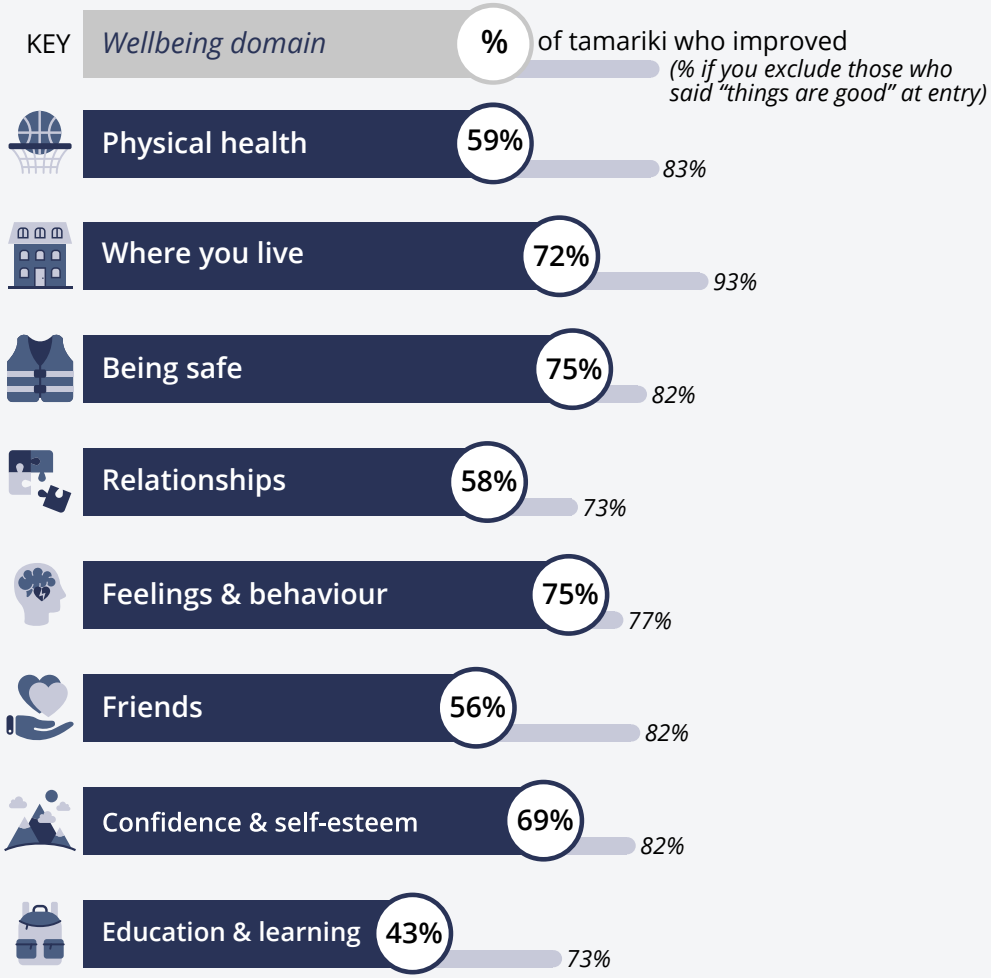
Measuring tamariki wellbeing

Every tamaiti who enters the Kōkihi ngā Rito programme begins a journey—one that they shape and direct themselves, supported by the dedicated mahi of NCIWR. Tamariki are encouraged to share their experiences, voice their priorities, and decide what goals matter most to them. Their insights guide the support they receive, ensuring it aligns with their needs and aspirations. Kaiārahi Tamariki use the My Star™ tool to facilitate meaningful conversations and reflections about tamariki wellbeing across eight key domains. This process helps tamariki identify their challenges, express progress in their own words, and guide the care they receive.

For this analysis, we focused specifically on My Star™ data collected at two points: when tamariki first enter the programme (intake) and as they prepare to leave (exit). We used paired t-tests—a statistical method designed to measure change within the same group over time—to evaluate the differences between intake and exit scores for 71 tamariki who participated in Kōkihi ngā Rito. This approach allowed us to quantify not just the size of the improvements but also confirm that these changes were real and unlikely to be due to chance. Effect sizes provided additional insight into the practical significance of the results, showing how substantial and meaningful the improvements were in the lives of these tamariki.

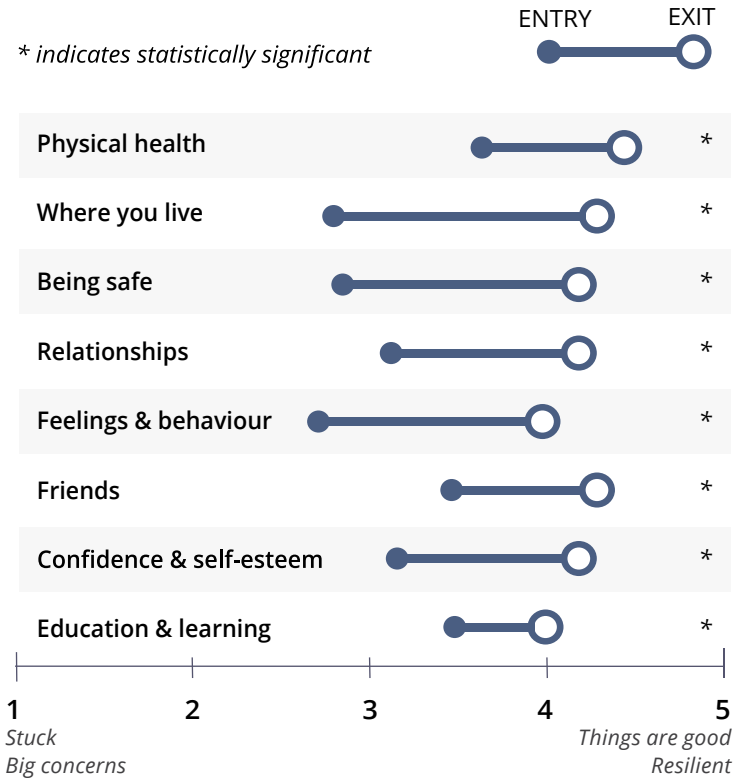
Percent of tamariki whose wellbeing improved

Split by My Star™ wellbeing domain



Average entry and exit scores

Split by My Star™ wellbeing domain



I just felt like [my Kaiārahi Tamariki] always understood my feelings... I knew I couldn't upset her.

- Tamaiti

The results are striking. Across all domains of wellbeing, tamariki experienced significant and meaningful improvements. Gains were particularly pronounced in the domains of *Being safe*, *Feelings and behaviour*, and *Confidence and self-esteem*, where Cohen's d quantified the effect sizes as "very large". The data show tangible progress for tamariki in Kōkihi ngā Rito, and mirrors the key findings from a qualitative evaluation conducted with tamariki interviews and focus groups [29]:

Being safe: Significant gains in this domain may reflect tamariki experiencing a greater sense of safety and stability by the end of the programme.

Confidence and self-esteem: Improvements in confidence show tamariki rebuilding their belief in themselves, trusting that they are worthy of safety and advocacy.

Feelings and behaviour: These gains may represent meaningful progress in children's ability to express, explore, and make sense of their feelings, especially the complex and often contradictory feelings they hold for their fathers.

These results were the same across ethnicity, gender, deprivation index, risk profile, and whānau composition, and statistical tests confirmed that no tamaiti attribute was a predictor of outcome. In other words, all tamariki saw similar outcomes trajectories, regardless of demographic factors.

Together, the numbers and stories tell the same truth: tamariki leave Kōkihi ngā Rito feeling safer and better equipped to cope, to heal, and to thrive. These changes offer a foundation for tamariki to move forward with greater safety, stability, and strength—a step toward brighter futures for themselves and their whānau.

Tamariki supported by Kōkihi ngā Rito experienced large, measurable, and statistically significant improvements across eight domains of wellbeing. These results are more than numbers—they reflect tamariki reclaiming safety, rebuilding their confidence, and rediscovering hope.

Comparing impact

The Kōkihi ngā Rito difference

Evaluation of My Star™ data showed that tamariki in the Kōkihi ngā Rito programme made measurable gains in their wellbeing, but we sought to understand whether these improvements were unique to the programme or part of a trend shared across other services. NCIWR provides support to thousands of tamariki each year, and while My Star™ is not routinely used across all services, a small subset of tamariki—just 43 children—had My Star™ assessments completed outside Kōkihi ngā Rito. We compared these My Star™ data to those from Kōkihi ngā Rito.

To account for the limitations in sampling and to ensure a robust comparison between the two groups, we approached the question from multiple different statistical perspectives. Our methods included paired t-tests, propensity score matching (PSM), and difference-in-differences (DiD), each offering different lenses through which to examine the quantitative data.

Across all approaches, the findings were clear and consistent: tamariki in Kōkihi ngā Rito experienced significantly greater improvements in key domains of wellbeing than children supported by other NCIWR services. The differences were particularly pronounced in *Feelings and behaviour, Relationships*, and *Confidence and self-esteem* areas of wellbeing.

Significance testing suggests these results are not a coincidence. Kōkihi ngā Rito was explicitly designed to create spaces where tamariki feel safe to explore emotions, rebuild trust, and rediscover their sense of self. The statistical results—consistent across methodologies—support the programme's promise: that when tamariki are supported to lead their own journeys, the changes are greater, deeper, and more enduring.

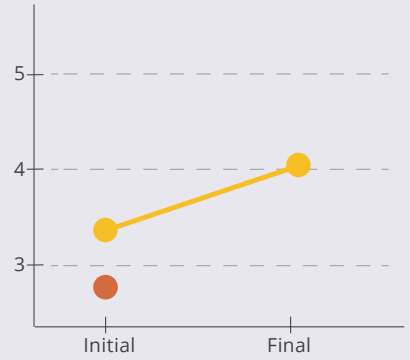
What makes these results even more compelling is the nature of the comparison group. We expect that the 43 tamariki outside Kōkihi ngā Rito were not a typical group; because Advocates deliberately chose to use the My Star™ tool with them, this group of tamariki likely reflected higher levels of need, risk, or circumstances requiring more intensive support—factors naturally aligning more closely with the Kōkihi ngā Rito cohort. The fact that Kōkihi ngā Rito tamariki still achieved greater gains underscores the programme's distinct and measurable impact.

While tamariki in other child services at NCIWR absolutely showed progress—a testament to the value of those supports—the gains for Kōkihi ngā Rito tamariki were consistently larger, their wellbeing trajectories steeper.

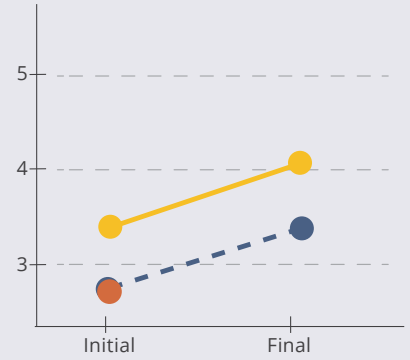
How to read a Difference-in-Differences (DiD) chart

*A Difference-in-Differences chart helps show the impact of an intervention (such as a programme or treatment) by comparing two groups over time: the **treatment group**, who experienced the intervention (in this case, the tamariki supported by Kōkihi ngā Rito) and the **comparison group**, who didn't (the tamariki supported by other services).*

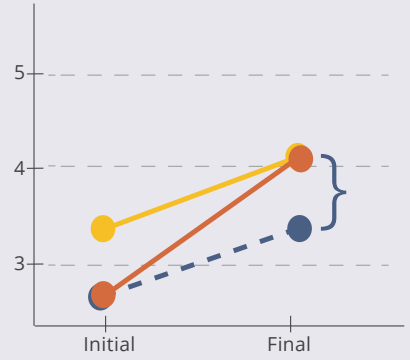
Example DiD chart



The yellow line on this chart shows the average change in wellbeing for the **comparison group**, between their initial and final My Star™ assessments. The orange dot shows the average initial score for the **treatment group**, tamariki supported by Kōkihi ngā Rito.



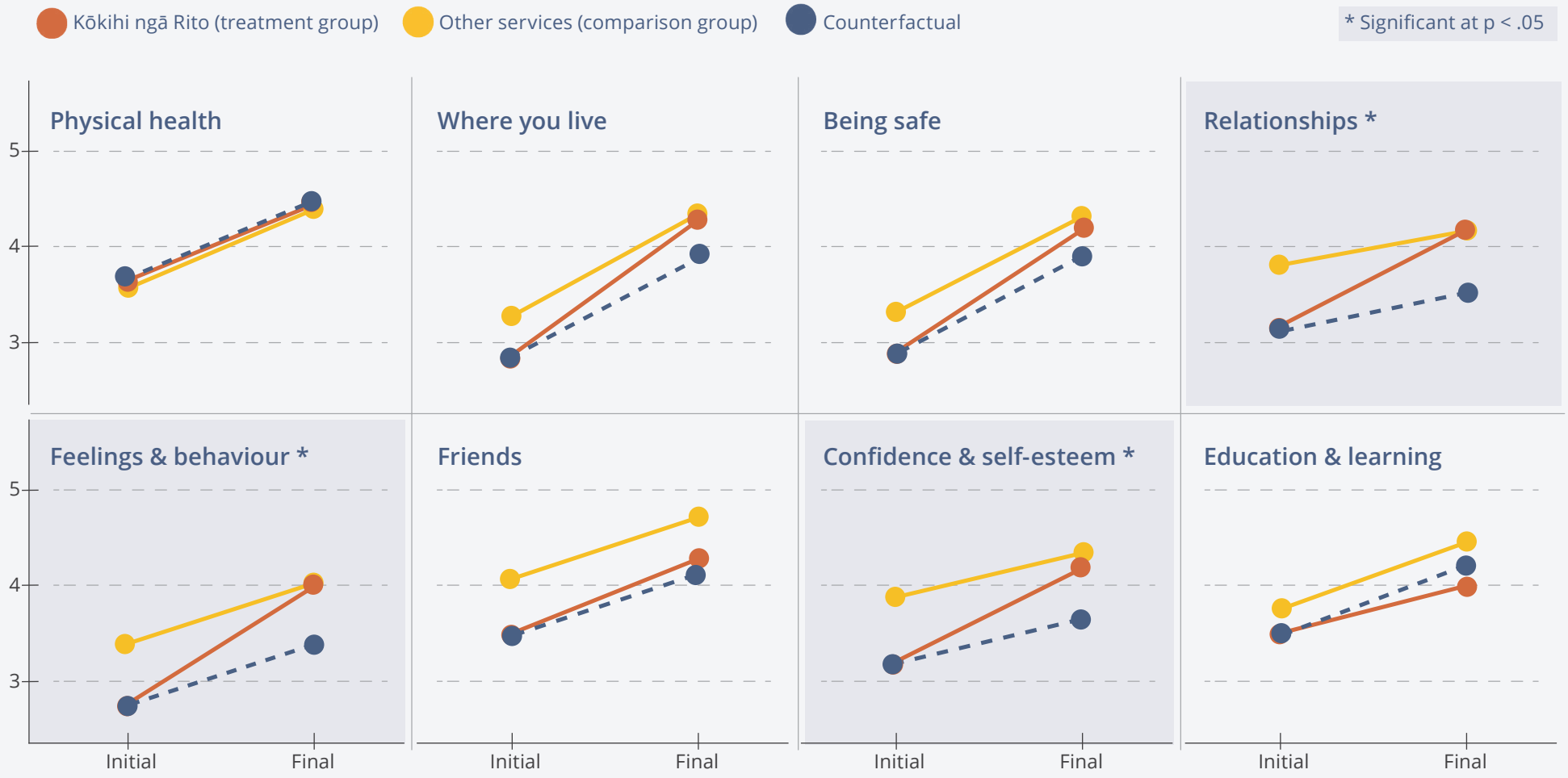
If the treatment has no effect, then we would expect to see the exact same trajectory of change in both groups, and the treatment group would look like the blue dotted line. This line is the **counterfactual**, or the change in wellbeing that we would expect to see for Kōkihi ngā Rito children if the programme has no influence on their wellbeing journey.



If the treatment does have an effect, then we expect to observe a different trajectory of wellbeing change in the treatment group, the orange line. The difference between where the counterfactual line ends (what we expected, if Kōkihi ngā Rito had no impact) and where the treatment group line ends (what we actually observed) is the "difference in differences" that gives this approach its name, and quantifies the impact of the intervention. Significant tests can then tell us whether this difference is likely to be meaningful or due to chance.

Comparing wellbeing change between Kōkihi ngā Rito and other services

Split by My Star™ wellbeing domain



Tamariki in Kōkihi ngā Rito experienced significantly greater improvements in wellbeing, particularly in *Feelings and behaviour*, *Relationships*, and *Confidence and self-esteem*, than children in other services. This difference is measurable, but it is also human—representing real lives changed and futures brightened.

Ripple effects

How child advocacy affects maternal wellbeing

Tamariki do not grow, suffer, or thrive in isolation. A child's wellbeing is deeply sensitive to their mother's wellbeing, and vice versa—when harm touches one, it affects the other[33,34,35]. This area of evaluation explored the other side of that relationship: can supporting tamariki also support their mothers? Using mothers' Empowerment Star™ data, we compared changes across nine domains of wellbeing for three groups of wāhine: the mothers of tamariki supported by Kōkihi ngā Rito, mothers of tamariki supported by other NCIWR services, and mothers whose tamariki received no direct support from NCIWR.

It is important to note that this lack of direct support stems from current funding limitations rather than any service delivery choice—NCIWR lacks the dedicated resources to provide comprehensive tamariki services across all refuges without specific programme funding. While this funding gap has inadvertently created a comparison group that strengthens these statistical analyses, we acknowledge the profound ethical discomfort this presents. Our commitment remains to eliminate these gaps entirely, even if it means weaker statistical comparisons in future evaluations.

Analysis revealed significant differences in wellbeing change between the three wāhine groups in two key wellbeing domains: *Children* and *Work and learning*. Deeper examination showed that when tamariki support from other NCIWR services, their mothers experienced significantly larger increases in wellbeing compared to mothers whose children had no direct support provided by NCIWR. Direct support for tamariki corresponded with measurable, statistically significant differences in their mothers' wellbeing.

For the Kōkihi ngā Rito cohort, results were less conclusive. Differences between this group and others did not reach statistical significance, which may be due to the small sample size—just 54 wāhine with complete, matched entry-exit pairs of wellbeing data, and whose tamariki had all been supported by Kōkihi ngā Rito—limiting our ability to detect subtle effects. Alternatively, it may reflect the greater harms experienced by tamariki in the Kōkihi ngā Rito cohort, which could naturally influence a mother's sense of wellbeing. Even with these limitations, the broader finding is significant: targeted support for tamariki has a meaningful impact on their mothers' wellbeing.

Advocating for the tamaiti is advocating for Mums, and vice versa.

- Kaiārahi Tamariki

Maternal wellbeing change and tamariki support

Split by Empowerment Star™ domain and wāhine cohort

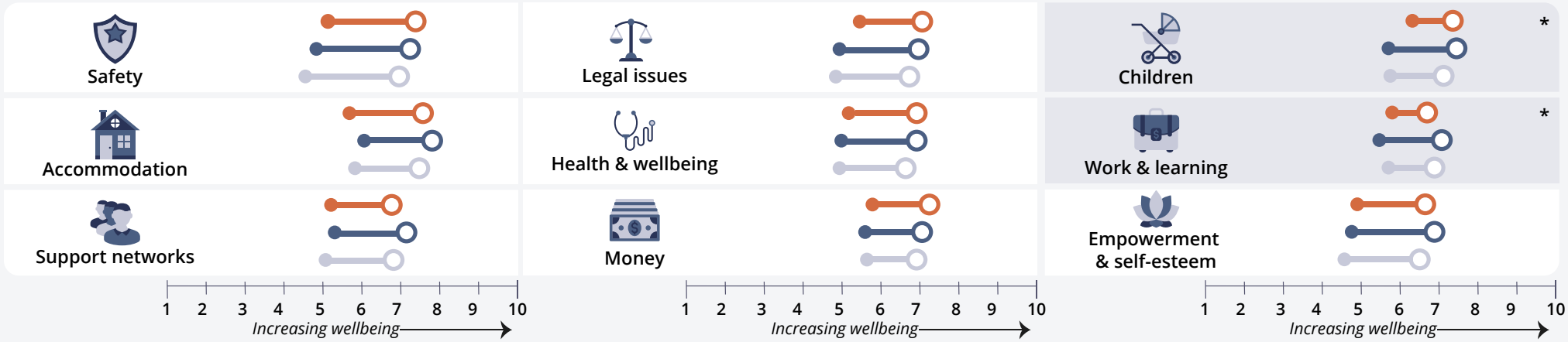
Each line represents the average wellbeing change during service

ENTRY — EXIT

** indicates statistically significant*

Colour indicates wāhine cohort, (support provided to her children)

- Kōkihi ngā Rito
- Other NCIWR service
- No NCIWR support



Wāhine cohort sizes

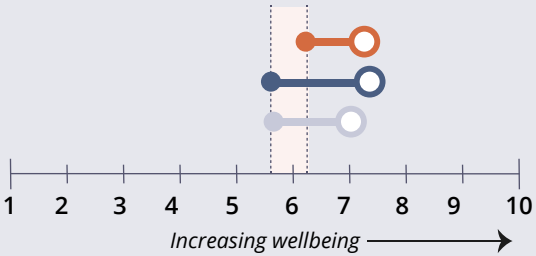
For every **100** mums supported by NCIWR:
2 had children in Kōkihi ngā Rito,
12 had children in another NCIWR service,
86 had children who received no support

At entry, wellbeing for Kōkihi ngā Rito mums is generally higher than that of other cohorts, in the domain of Children

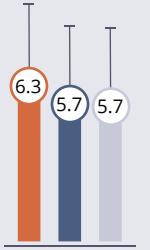
Focusing solely on the changes occurring during service delivery can sometimes overlook other important narratives. For instance, while the wellbeing improvements reported by Kōkihi ngā Rito mothers were smaller compared to those in other NCIWR services, this is likely due to differences in their starting points.

When it comes to their tamariki, mothers of children in Kōkihi ngā Rito already report significantly higher wellbeing around their children at the beginning of service—suggesting that simply knowing their child is supported may positively impact a mother's own wellbeing.

Wellbeing change, *Children* domain



Initial wellbeing, *Children* domain



This result is statistically significant at a 10% level, but not at the stricter 5% level. The unequal sizes of the wāhine cohorts may influence significance tests, and further exploration of these patterns could be informative.

In the *Work and learning* domain, the ripple effect is particularly intriguing. While the connection may seem less direct, practical support provided through child advocacy—such as tamariki being picked up from school or supported with after-school activities—could perhaps ease pressures on mothers. This could create space for wāhine to reengage with work, education, or other opportunities that support their own aspirations and stability.

The patterns seen in this data suggest a deeper relationship worth exploring. Thus far, qualitative evaluation has predominantly focused on the experiences of tamariki and mothers' perceptions of their child's journeys. Additional interviews with wāhine across multiple cohorts could provide valuable insight into how these dynamics unfold in practice and illuminate new ways to better serve mothers experiencing family violence.

Sometimes for that hour and half [while the Kaiārahi Tamariki had my son], it would mean I could just sort of breathe, maybe do something very small for me that gave me that extra, so that when I picked him up from school I was one hundred percent.

- Kōkihi ngā Rito Mum

The *Children* domain highlights another important truth: when tamariki are seen, heard, and supported, their mothers' wellbeing reflects this shift. Gains in this area speak to the interconnected nature of family wellbeing—when children are safer, more confident, and supported, their mothers feel it too.

Although the programme has not yet achieved enough data sampling for the statistical power required to fully tease apart the specific impact of Kōkihi ngā Rito compared to other NCIWR services, the broader findings remain significant and point to valuable avenues for future exploration and elaboration.

Support for the child victims of family violence corresponds with measurable gains for their mothers, affirming the interdependence of tamariki and maternal wellbeing. When tamariki grow stronger, safer, and more secure, their mothers may also gain opportunities to heal, rebuild, and thrive—strengthening the whānau as a whole. When children thrive, their whānau are lifted too.

A rising tide

How Kōkihi ngā Rito changes the system

The Kōkihi ngā Rito programme was designed to deliver a new standard of care for tamariki—intentional, sustained, and deeply child-centred. It sets itself apart from other services within NCIWR by prioritising tamariki wellbeing through unique service delivery elements. To understand these distinctions, we evaluated three key measures: the length of referrals, the hours of direct support provided, and the depth of documentation through case notes.

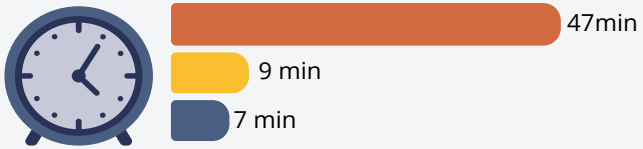
Tamariki were grouped into three cohorts: those supported by Kōkihi ngā Rito, those supported by other NCIWR services at refuges that deliver the programme, and those at refuges without it. An additional analysis focused on high-risk tamariki—children whose mothers reported all six child-related risks—to ensure that potential differences in harm were not confounding any observed differences between tamariki cohorts.

The findings were striking. Across all service delivery measures, tamariki in Kōkihi ngā Rito received significantly more care—longer referral durations, vastly more hours of direct support, and more detailed case notes. Across all cohort comparisons, these differences were statistically significant, suggesting that that Kōkihi ngā Rito is not only distinct in its design but also in its delivery. By these measures, the practical implementation of the programme adheres closely to the theoretical underpinnings of the service model.

Average referral duration

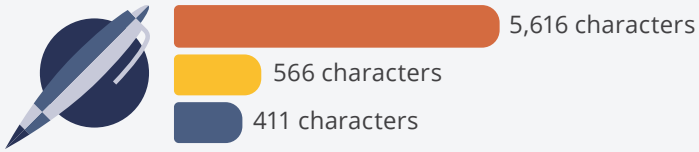
Referrals were 2x longer.
Tamariki in Kōkihi ngā Rito remained in service for over six months on average—more than double the time for children in other services. This extended timeframe reflects the programme's commitment to providing sustained, tailored care to meet tamariki where they are.

Average daily support hours



Tamariki received 5-7x more hours of support.
Even corrected to account for longer referral durations, Kōkihi ngā Rito tamariki received 5x the direct support hours of those in other services at the same refuge and over 7x more than those at refuges without the programme. These additional hours create space—to listen, to build trust, and to provide care that responds to tamariki.

Average case note length



Case notes were 9-13x longer.
Client files for Kōkihi ngā Rito tamariki averaged more than 5,000 characters per case note, roughly 1.5–2 pages of single-spaced text. These case notes, 9x longer than those for children in other services at the same refuge and 13x longer than those at other refuges, document the care and attention dedicated to each child's journey.

Colour indicates tamariki cohort (who supported this child?)

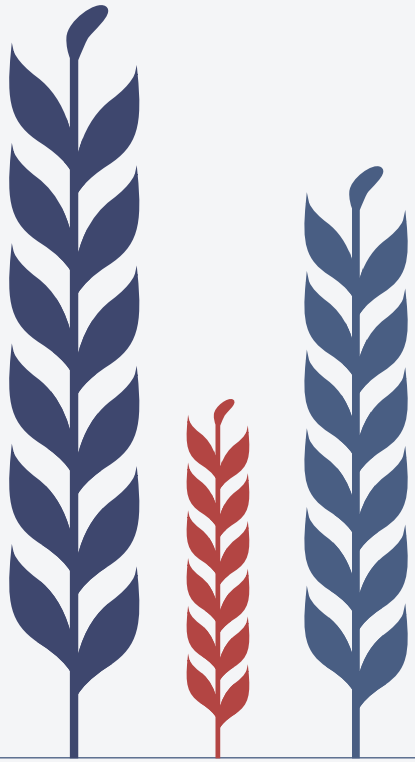
- Kōkihi ngā Rito
- Other service (at a refuge that delivers Kōkihi ngā Rito)
- Refuge without Kōkihi ngā Rito

It's the time to build the rapport, to hear their voice from their mouth, not what their parents think, or their grandma, or their teacher, or [someone] they had one session with.

- Kaiārahi Tamariki

Beyond the programme's direct participants, its impact extended across entire refuges. Tamariki who were not enrolled in Kōkihi ngā Rito but were supported by other services at a refuge delivering the programme received **1.3x more support hours** and **1.4x longer case notes** than those at refuges without it. For children with higher risk profiles, these differences were even more pronounced, suggesting that the presence of Kaiārahi Tamariki could uplift service delivery across all refuge operations.

This "rising tide" effect demonstrates the programme's broader influence. While the exact mechanisms are unclear, the specialised training, intentional practices, and high standards set by Kaiārahi Tamariki could inspire their colleagues—even those delivering different services—to engage more deeply with tamariki and document their care more comprehensively. Further qualitative research and exploration with kaimahi interviews could uncover the drivers of this change and amplify its impact.



It's a different way of working and it takes time to realise why writing notes is important, and just how everything you do is advocacy. It's more involved than other roles I've had, there is more flexibility—but the more you know you can do, the more you do, and it keeps on going.

- Kaiārahi Tamariki

The quantitative evidence underscores a genuine difference in service delivery, evidence of the programme's fidelity to its theoretical framework. Kōkihi ngā Rito is not just distinct in its vision—it delivers care that is measurably different, setting a new benchmark for child-centred services. When we invest in tamariki, we may do more than change individual lives—we can transform the systems that support them, lifting everyone closer to the care and safety they deserve.

The Kōkihi ngā Rito programme delivers measurably different care—longer, deeper, and more detailed support that reflects its intentional, child-centred design. The presence of specialist Kaiārahi Tamariki at a refuge appears to uplift service delivery across the entire refuge, suggesting a rising tide effect that raises the standard of care for all tamariki.

Case notes as evidence

Capturing the language of advocacy

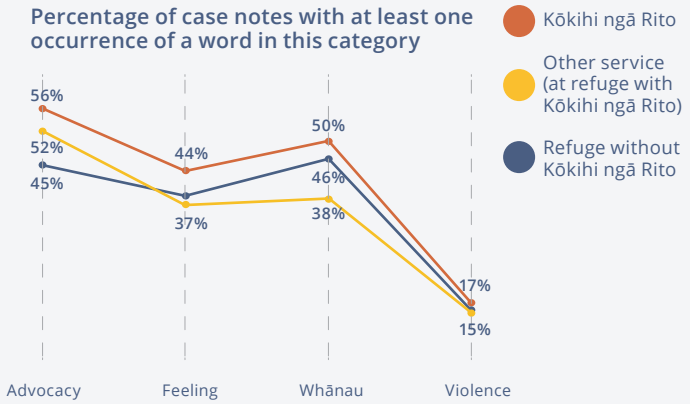
Case notes are more than just administrative records. They are a testament to the relationships forged between tamariki and those who support them—snapshots of trust, of breakthrough moments, of quiet victories. Our earlier analyses showed that Kōkihi ngā Rito case notes are significantly longer than for other services or refuges, so this final analysis dove into the content of that documentation to understand whether the distinctive approach of Kōkihi ngā Rito is reflected in the words used by Kaiārahi Tamariki to document care. Specifically, we compared the words, content and themes of case notes across three cohorts: tamariki supported by Kōkihi ngā Rito, tamariki supported by other services within refuges delivering Kōkihi ngā Rito, and tamariki supported at refuges without the programme.

The case notes of tamariki supported by other services were largely similar to one another, but in multiple ways, the case notes of Kōkihi ngā Rito were distinct.

I don't think the depth of their true experiences, or the true force of the child would have been able to be captured without this pilot.

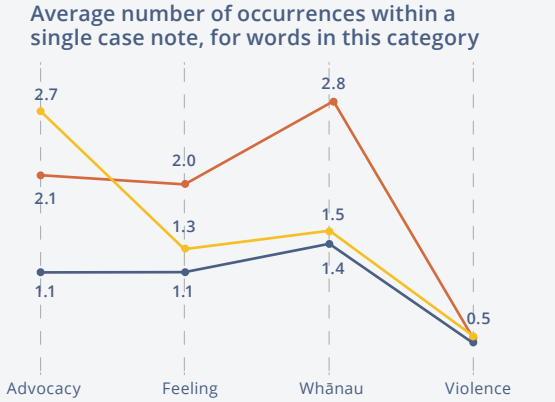
- NCIWR Manager

More Kōkihi ngā Rito case notes mentioned whānau, feelings and advocacy.
Individual Kōkihi ngā Rito case notes were slightly (but significantly) more likely to include at least one word about whānau, feelings, and advocacy. At first glance, the comparison at this initial level of inquiry suggests relatively minor differences between services; however, certain words accounted for large components of the similarity. Most children's case notes mention "mum" (whānau), "safety" in the context of safety planning (advocacy), and often describe the child's emotional state (feeling).



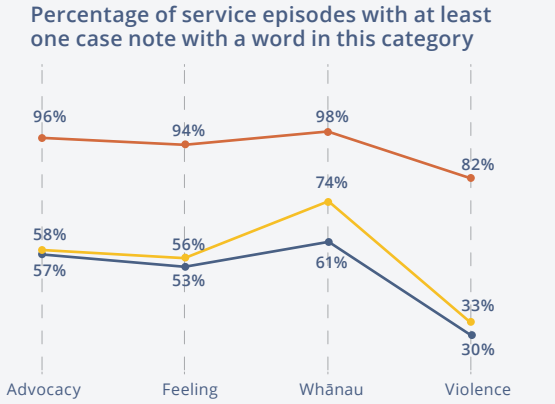
Kōkihi ngā Rito case notes included deeper discussions of whānau and feelings.

Kōkihi ngā Rito case notes included, on average, 2x as many occurrences of feeling and whānau words per note. In other words, when a Kōkihi ngā Rito case note spoke of feelings or whānau, it was rarely just a brief mention; the notes tended to speak more about that subject, providing more detailed discussion with multiple instances of the word repeated in the note.



Kōkihi ngā Rito case files were much more likely to include descriptions of feelings, whānau, advocacy, and violence.

Because Kōkihi ngā Rito case files include so many more individual case notes than other services, it was also important to consolidate and compare those complete tamariki files. More than 90% of all case files for Kōkihi ngā Rito tamariki included case notes discussing feelings, advocacy and whānau—compared to just 53-74% of the casefiles for tamariki in other services. Discussions of violence were also over 2.5x more frequent within Kōkihi ngā Rito case files than in those of children supported elsewhere.



Kōkihi ngā Rito case notes more faithfully recounted a child's words.

Most noticeable when discussing violence and feelings, Kōkihi ngā Rito case notes were markedly different in language, more closely reflecting a child's voice with words like fight, hide, punch, kill, hurt, stab—whereas case notes for tamariki in other services were more likely to use clinical language to describe similar experiences. The words "worried" and "scared" appeared over 2x as often in Kōkihi ngā Rito case notes, describing tamariki feelings with as much care as the more standard descriptions of "abuse", "safety", and "violence".

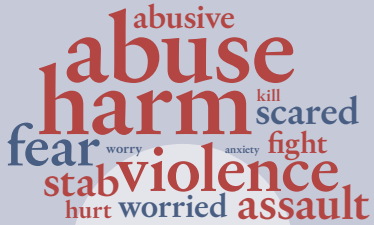
Case notes are often overlooked as a measure of service quality, but they offer a unique lens into how care is provided. The detail and emotional depth observed in Kōkihi ngā Rito case notes provide further evidence that this programme delivers a distinct level of support. Its work goes beyond practical tasks to provide advocacy attuned to the emotional lives of child, which sees, hears, and holds space for their experiences and reflects those authentically even in official documentation.

The size of the circles represents how many case notes the average service episode contains, which is important context when interpreting percentages. Although "worried" is only 2.15x more likely to appear in a Kōkihi ngā Rito case note compared to one from a refuge that does not offer the programme (4.8% ÷ 2.2%), because there are so many more case notes on an average Kōkihi ngā Rito service episode (38.3 vs 1.9), the word "worried" is 43.4x more likely to appear in at least one case note on a tamāiti file in Kōkihi ngā Rito:
(4.8% x 38.3 = 1.838 notes/service episode)
vs (2.2% x 1.9notes = 0.042 notes/service episode)
= 1.838 ÷ 0.042 = 43.4

The size of each word represents the percent of case notes in which it appears, within each cohort of tamariki. For example, the word "worried" appears in 4.8% of Kōkihi ngā Rito case notes, 1.5% of those for non-Kōkihi ngā Rito tamariki at a refuge that offers the programme, and 2.2% of case notes for tamariki at refuges that don't offer Kōkihi ngā Rito.

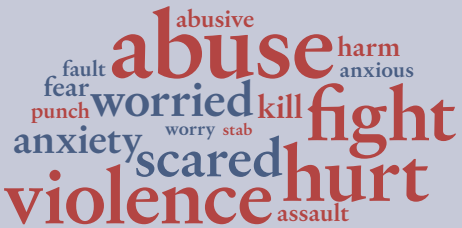
38.3 case notes per service episode

Kōkihi ngā Rito



6.7 case notes per service episode

Other service (at a refuge that delivers Kōkihi ngā Rito)



1.9 case notes per service episode

Refuge without Kōkihi ngā Rito

Kōkihi ngā Rito case notes talk about tamariki feelings and whānau much more often than those of other services, and when they discuss these subjects, they do so in more detail; further, descriptions of violence are most often documented in a child's words, reflecting the focus on encouraging tamariki to articulate their own feelings, experiences, and needs.

Risk and reach

Mapping harm and possibility

When wāhine enter NCIWR services, advocates help them to complete a risk assessment designed to identify immediate and ongoing safety concerns. While this process focuses primarily on their own experiences of family violence, it also includes specific questions about risks and harms affecting their tamariki. These responses reflect the mothers' perceptions of what their children have experienced, offering an indirect window into the specific harms tamariki suffer in homes marked by family violence.

The data collected through risk assessments paints a vivid and often troubling picture of that abuse. Between January 2018 and July 2024, over 20,000 risk assessments were completed by wāhine entering service with NCIWR. Each of these assessments is more than a statistic; they represent mothers recounting the harm their children have faced—an inventory of threats, separations, physical harm, and exposure to multiple types of abuse.

The assessment questions captured six key types of child-related harm:

- Children threatened with harm
- Children taken or kept from mum
- Children physically harmed
- Mum verbally abused in front of children
- Mum physically harmed in front of children
- Mum forced to do something sexual in front of children

What is “risk”?

The concept of “risk” appears in any discussion of family violence, but NCIWR adopts a nuanced and dynamic approach. Recognising that risk is not a static condition but a fluctuating interplay of factors, they view risk assessment as an ongoing, relational process rather than a one-off task[28]. This perspective is informed by their extensive frontline experience and the lived realities of wāhine and tamariki navigating family violence.

At the heart of their approach is the belief that those experiencing violence are the most knowledgeable about their own safety. NCIWR prioritises building trust and listening deeply to the perspectives of wāhine and tamariki, integrating their insights into comprehensive risk evaluations. These evaluations consider not only immediate physical danger but also psychological, social, and cultural dimensions of harm.

NCIWR emphasises the importance of shifting away from a purely quantitative model of risk assessment to one that values qualitative, context-rich insights. They advocate for a framework that centres the voices of those impacted by violence, enabling tailored and adaptive safety plans. These comprehensive risk assessments also form a critical component of the evaluation presented in this report, offering invaluable and otherwise impossible insights into the experiences of wāhine and tamariki navigating family violence.

The findings are stark: 71% of all wāhine reported at least one child-related harm, with almost one in three revealing their children had experienced all six types of abuse. These are not abstract figures—they represent tamariki threatened with harm, physically abused, separated from their mothers, and forced to witness verbal, physical, and even sexual abuse. From the roughly 15,000 wāhine supported by NCIWR every year, these figures extrapolate conservatively into over 25,000 tamariki annually who are harmed by family violence. Of these victims, currently fewer than 1 in 5 receive support from NCIWR, and less than 1% participate in the Kōkihi ngā Rito programme.

This analysis underscores both the widespread prevalence of harm to children and the immense opportunity for programmes such as Kōkihi ngā Rito to reach tamariki in need, particularly given that the ready availability of identifiable tamariki details and risk profiles could facilitate access, engagement, and prioritisation. These children are known to NCIWR, but despite the harm they have experienced, most receive no direct support to help them heal.

By supporting the child victims of family violence through targeted advocacy, the programme has the potential to help thousands of children and whānau across Aotearoa to rewrite their own narratives.

[Without Kōkihi ngā Rito], tamariki would still be silent and it would mean that for the next generation, there will be no generational change because that cycle will keep going.

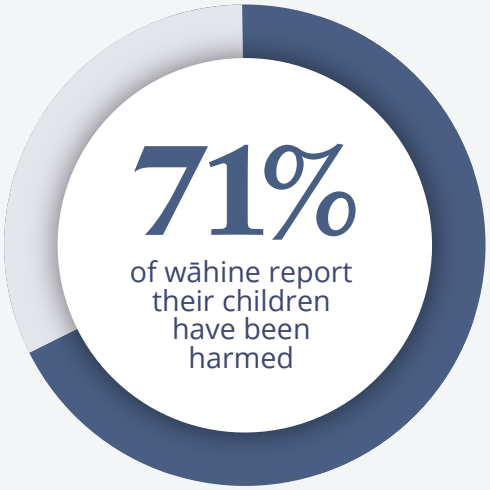
-NCIWR Refuge Manager

I think [without Kōkihi ngā Rito, my daughter] would be stuck in the hole, how she was at the start... just in her shell, feeling no self-worth. With domestic violence, it's hard to explain what it does to you as a person.

- Mum of Kōkihi ngā Rito child

Understanding the prevalence of child-related risk and harm

Insights from mothers' disclosures



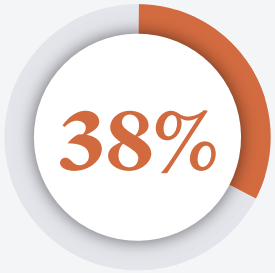
This equates to an estimated

25k tamariki annually

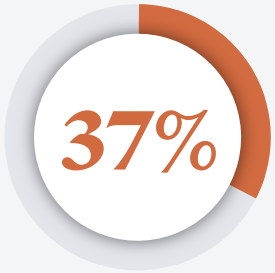
The vast majority receive

no support*

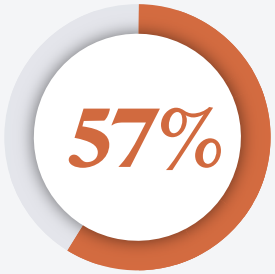
Tamariki have been threatened with harm



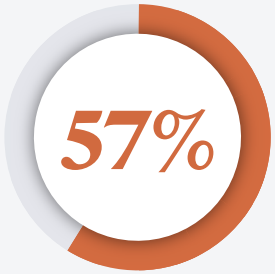
Tamariki have been physically harmed



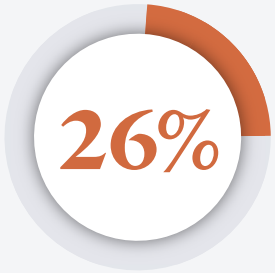
Tamariki have been taken or kept from mum



Mum has been physically abused in front of the children



Mum has been verbally abused in front of the children



Mum has been forced to do something sexual in front of the children

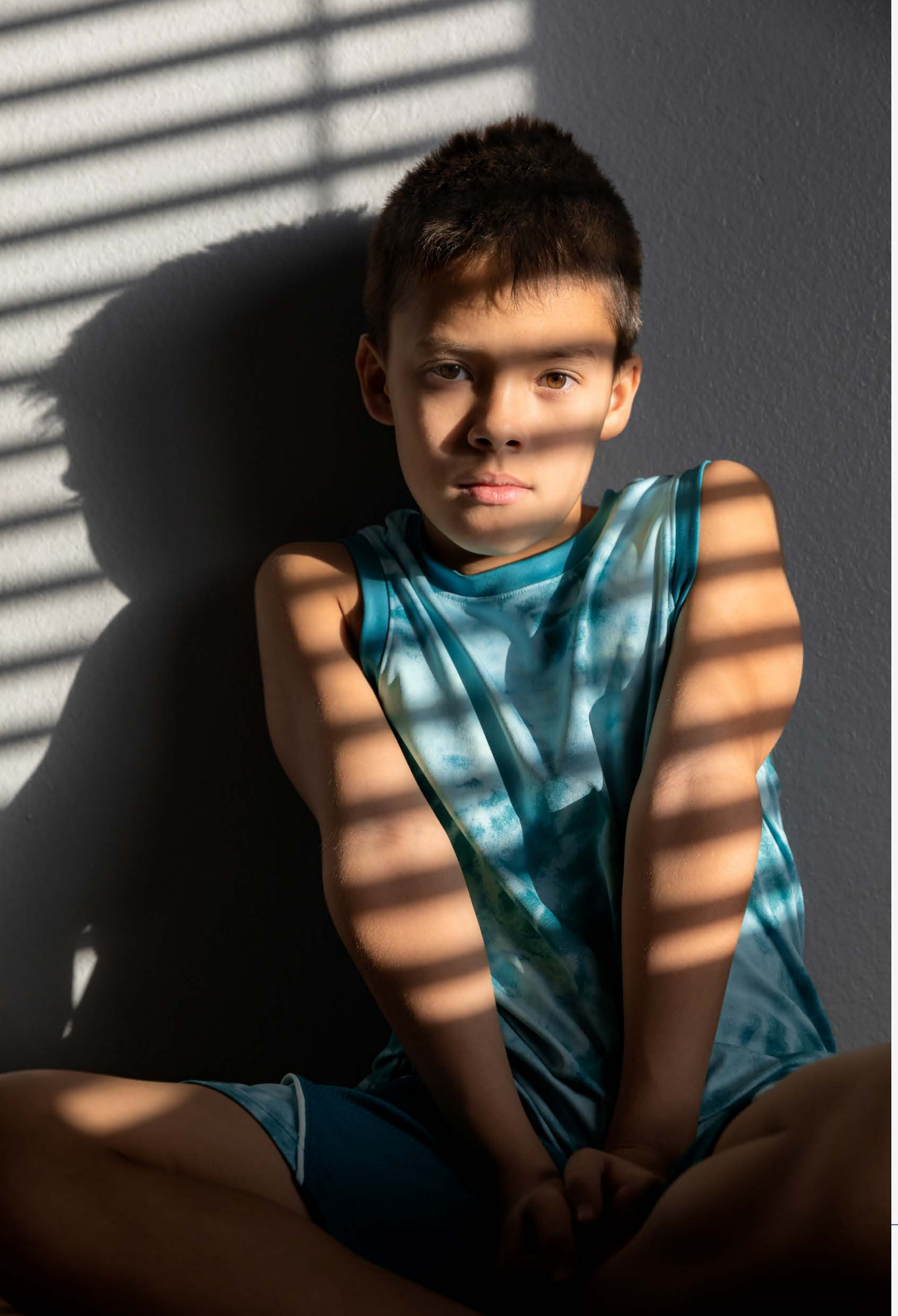
I was traumatised by what my dad did.

- Boy, 11 years old

Their whole world has been turned upside down.

- Mum of Kōkihi ngā Rito child

*no support from NCIWR (other social services may be involved, but there is no knowledge of any other supports provided to these children within NCIWR data



Risk assessments indicate conservatively that the mothers of over 25,000 tamariki are supported by NCIWR each year. Of these children, one in three has been threatened with harm, physically abused, separated from their mother, and forced to witness verbal, physical, and even sexual abuse.

The vast majority of these child victims receive little or no targeted support.

Discussion

Discussion

How statistics become solutions

The findings of this quantitative evaluation highlight the profound impact of Kōkihi ngā Rito, demonstrating statistically significant improvements in tamariki wellbeing across multiple domains. The My Star™ assessments indicate that participation in the programme is associated with large and meaningful gains, with effect sizes surpassing those observed in other service models. This reinforces the strength of Kōkihi ngā Rito's child-centred, relational advocacy model, which ensures tamariki are seen, heard, and supported in ways that extend beyond traditional family violence services.

Beyond individual outcomes, the data also suggest systemic benefits. Tamariki in Kōkihi ngā Rito receive longer and deeper engagement, with more time in care, higher levels of direct support, and richer documentation in case notes. Interestingly, these effects extend beyond the programme itself, influencing the broader refuge environment. Even tamariki not formally enrolled in Kōkihi ngā Rito receive more hours of support per day and more detailed documentation when they are in a refuge that offers the programme. This ripple effect raises important questions about staff training, practice shifts, and the potential for wider cultural change within the sector.

The results of this evaluation highlight the effectiveness of the Kōkihi ngā Rito programme, but like all research, there are limitations to what the data can tell us. Understanding these limitations helps ensure that conclusions are drawn carefully and that future evaluations continue to refine and build upon this work.

Understanding the limitations

While the evidence supports the effectiveness of Kōkihi ngā Rito, we raise several limitations to consider when interpreting the results.

Statistics tell us about patterns, not certainties

When we say a result is "statistically significant," we mean that a strong pattern has emerged within the data and the likelihood of this pattern occurring by chance is very low. However, statistical significance is about probability, not certainty—there is always some degree of uncertainty. In this evaluation, the consistency of findings across multiple statistical approaches strengthens confidence that the observed results are less likely to be due to random variation.

Even when we see a strong relationship between participation in Kōkihi ngā Rito and improved wellbeing, we cannot assume direct causation. Other factors—such as changes in a child's home environment, school life, or additional support they may receive elsewhere—could also play a role. In any evaluation, results should be considered alongside expert knowledge to form well-rounded conclusions. Importantly, the fact that Kōkihi ngā Rito is specifically designed to improve the same wellbeing areas where tamariki saw the greatest gains adds further weight to the relationship.

Cultural context and interpretation is important

This evaluation used statistical methods that, while robust within Western frameworks, may not fully capture or interpret outcomes through te ao Māori. Although the research team included Māori expertise, the unique cultural contexts and worldviews that inform Māori understandings of wellbeing were not specifically centered in this analysis and future work should expand on this foundation to integrate kaupapa Māori frameworks to better represent the cultural dimensions of wellbeing for tamariki Māori.

Real world data is not a controlled experiment

This quantitative evaluation uses operational data—information recorded day-to-day for the purpose of delivering care, rather than data collected in a controlled and randomised research setting. This means natural variations exist in how that data is recorded across different refuges, staff, and situations. For example, some workers write more detailed case notes than others, and different refuges may have slightly different approaches to service delivery.

Unlike a scientific trial, where participants are randomly assigned to different groups, tamariki enter Kōkihi ngā Rito based on their individual needs. This means we cannot completely separate the impact of the programme from other life circumstances, and we cannot apportion children to cohorts to ensure an even distribution of age, gender, ethnicity, need, and other factors. Although demographic variables were tested for inclusion in PSM and DiD modelling, we can only attempt to account for these factors post-hoc, rather than controlling them by design.

Every tamaiti is an individual

Every child's experience of family violence is unique and complex, influenced by whānau, community, and personal factors that we can't always measure. While data gives us valuable insights, it cannot tell us everything about a child's emotional world, resilience, or long-term healing journey. This quantitative evaluation should be considered hand in hand with qualitative input from tamariki.

What these findings mean

Despite these limitations, the results are clear and compelling—Kōkihi ngā Rito is providing a high level of support and advocacy that is making a real difference in the lives of tamariki. At the same time, the data highlight a major gap: while thousands of tamariki are affected by family violence every year, only a small fraction receive the depth of care that Kōkihi ngā Rito provides. The programme's success shows that specialised, child-focused support works, but it also raises an urgent question:

How do we ensure that more tamariki receive this level of care?

The next steps are not just about scaling up Kōkihi ngā Rito, but also about deepening our understanding of what works best for tamariki, strengthening integration across refuges, advocating for policy and funding support, and ongoing learning to refine and expand impact.

Expanding Kōkihi ngā Rito to reach more tamariki is a clear priority. Increasing the number of Kaiārahi Tamariki across additional refuges would extend the programme's benefits to a larger cohort of children, particularly in regions with the highest child-related risk. Critical to this expansion is maintaining fidelity to the service model that produced these positive outcomes—including the comprehensive training, and supervision structures that define Kōkihi ngā Rito. Developing targeted outreach strategies and referral pathways would ensure that those most at risk are identified early and connected to support as soon as possible. Given the programme's demonstrated success, scaling its reach is essential to reducing the gap between the number of tamariki experiencing harm and those receiving specialised care.

Beyond programme expansion, the evaluation suggests that Kōkihi ngā Rito's child-centred approach has a wider positive influence across refuges. Even tamariki not formally enrolled in the programme received more hours of support and had more detailed case notes when in refuges that offer Kōkihi ngā Rito. This ripple effect presents an opportunity to embed its best practices into standard refuge operations. Providing training and resources to all refuge staff would help integrate child-centred, relational advocacy into other tamariki services, improving the consistency and quality of care. Strengthening knowledge-sharing and collaboration between Kaiārahi Tamariki and other staff could further support the spread of effective child advocacy approaches and identify additional levers and mechanisms of influence.

Addressing the systemic scale of unmet need requires not only direct service expansion but also long-term investment in the sector. Securing policy and funding support from government agencies, funders, and community stakeholders will be critical to ensuring that child-focused interventions remain sustainable. Increasing investment in staffing, training, and infrastructure will help bridge the gap between service demand and service availability, ensuring more tamariki receive the care they need.

The evaluation also raises important questions about the interconnectedness of tamariki and maternal wellbeing. The data suggest that supporting tamariki has positive flow-on effects for their mothers, particularly in areas related to parenting and economic stability. Future research should further explore this relationship through mixed-methods studies that incorporate maternal voices, providing a deeper understanding of how child-centred interventions contribute to overall whānau wellbeing.

Beyond service delivery, there is an urgent need to advocate for systemic change. The findings highlight the importance of culturally safe, child-centred approaches in the family violence sector. Raising awareness of the impact of Kōkihi ngā Rito and advocating for wider adoption of its principles could strengthen child advocacy across the sector. This includes embedding best practices from the programme into broader family violence interventions to ensure tamariki voices remain at the centre of all support services.

Engagement with communities and whānau is also essential to refining and strengthening the programme. Involving tamariki, their caregivers, and refuge staff in co-designing programme improvements will ensure that future iterations of Kōkihi ngā Rito remain responsive to the realities and needs of those it serves. Expanding whānau-based approaches that recognise the interconnected nature of healing could also enhance long-term outcomes for both tamariki and their families.

We knew [child advocacy] was a missing link for years and we were trying to work out how to do that. That was our missing link. Now our chain is complete. The biggest difference that I see from a year ago to now is the commitment and the drive. I know now that service is going to be delivered the way that we have all dreamed and envisioned, not having sleepless nights thinking, "what are those children actually going to learn?"

- NCIWR Manager

Final conclusions

The results of this evaluation provide robust evidence that Kōkihi ngā Rito is delivering transformative support to tamariki affected by family violence. The programme's child-centred, relational approach leads to stronger, more meaningful engagement, with wellbeing gains that appear to extend beyond individual children to influence their mothers, whānau, and the wider refuge environment.

However, these findings also highlight an urgent need for action. Despite the success of Kōkihi ngā Rito, only a fraction of tamariki receive this level of care. Tens of thousands of tamariki, the child victims of family violence, remain invisible in our response systems—their trauma unaddressed, their voices unheard, and their potential diminished as they navigate a critical developmental window without access to the specialised, child-centred support that could help them.

The challenge ahead is not just about scaling a successful programme—it is about redefining the standard of care for tamariki in the family violence sector. The effectiveness of Kōkihi ngā Rito demonstrates that when we invest in children, we invest in futures, in families, and in lasting social change. The next step is ensuring that every tamaiti affected by family violence has access to the support they need to heal, thrive, and reclaim the future they deserve.

The human rights of the child have to be heard and be validated and be seen.

- NCIWR Manager

Technical Results

Outcomes Star™ My Star™ pre-post analysis:
Tamariki supported by Kōkihi ngā Rito

Statistical methods

A series of paired t-tests were conducted to assess the impact of the Kōkihi ngā Rito programme on tamaiti wellbeing, which was measured across eight domains using the Outcomes Star™ My Star™ tool. This tool uses a 1-5 scale to measure wellbeing in eight domains, with readings collected at intake to the programme (initial) and before discharge (final) for each child (n=71). Paired t-tests were selected as the most appropriate analytical method as they specifically examine within-subject changes over time while controlling for individual differences. This makes them particularly sensitive to detecting intervention effects.

The paired t-test's assumption of normally distributed differences between pre- and post-treatment outcomes is well-supported by simulation studies, which demonstrate the test's robustness to non-normality when sample sizes exceed 30, in accordance with the Central Limit Theorem[36,37,38,39]. Since extreme values can substantially influence means and test results, we carefully screened for outliers prior to analysis. Our examination of QQ plots revealed slight underdispersion; however, given our sample size of 71, this departure from normality was deemed acceptable for proceeding with parametric testing. We used Cohen's d to calculate effect sizes and quantify the magnitude of treatment-related changes.

Interpretation of findings

As shown in Table 1, the results suggest that participation in the Kōkihi ngā Rito programme had a significant and substantial effect on the wellbeing of tamariki across all eight domains measured. Using Cohen's conventional thresholds (0.2 small, 0.5 medium, 0.8 large)[40], the observed effect sizes ranged from medium to very large ($d > 1.0$). The strongest improvements were seen in *Friends and behaviour*, followed by *Where you live* and *Being safe*. Even in *Education and learning*, which showed the smallest effect size, the improvement was still substantial. All changes were positive and statistically significant at $p<0.001$, indicating consistent improvement across all domains of wellbeing.

Limitations

The paired t-test only captures within-pair differences and does not control for confounding variables that may influence the outcome. We acknowledge that the experience of family violence can be a profoundly chaotic and deeply individualised journey, each with its own set of external factors and influences. Additional limitations include the reliance on self-reported data, which may be subject to response bias, and the absence of a control group, which limits our ability to isolate programme effects from natural maturation or other external influences. The varying duration of programme participation may also impact outcomes. Additionally, the interpretation of a significant result in a paired t-test is limited to the difference in means rather than the full distribution of change and does not imply causation or provide insight into the reasons behind observed differences.

Conclusions

The paired t-test analyses provide robust evidence that participation in the Kōkihi ngā Rito programme is associated with significant improvements in tamaiti wellbeing across all eight My Star™ domains. The consistency of positive changes across domains, coupled with the magnitude of the effect sizes and strong statistical significance, suggests that the programme's holistic approach effectively supports multiple aspects of child wellbeing. Future research might explore the longer-term sustainability of these improvements and examine which specific programme elements contribute most strongly to positive outcomes.

Table 1. Paired t-test results for Kōkihi ngā Rito tamariki (n=71)

My Star™ Domain	Average initial	Average final	Average change	Standard deviation	SEM	t	p	95% confidence interval	Cohen's d	Effect size
Physical health	3.63	4.42	0.79	1.21	0.14	5.51	<0.0001*	(0.50 - 1.07)	0.77	Medium-to-large
Where you live	2.82	4.27	1.45	1.46	0.17	8.36	<0.0001*	(1.10 - 1.80)	1.28	Large-to-very large
Being safe	2.87	4.17	1.30	1.40	0.17	7.81	<0.0001*	(0.96 - 1.63)	1.21	Large-to-very large
Relationships	3.14	4.17	1.03	1.37	0.16	6.31	<0.0001*	(0.70 - 1.35)	0.94	Large
Feelings and behaviour	2.73	3.97	1.24	1.18	0.14	8.88	<0.0001*	(0.96 - 1.52)	1.44	Very large
Friends	3.46	4.27	0.80	1.37	0.16	4.94	<0.0001*	(0.48 - 1.13)	0.75	Medium-to-large
Confidence and self-esteem	3.17	4.17	1.00	1.22	0.14	6.91	<0.0001*	(0.71 - 1.29)	1.10	Large
Education and learning	3.48	3.99	0.51	1.18	0.14	3.62	0.0006*	(0.23 - 0.79)	0.46	Small-to-medium

A guide to interpretation:

Average initial

This is the mean of the entry assessment readings for each domain. For example, in *Physical health*, the average entry reading for Kōkihi ngā Rito tamariki is 3.63.

Average final

This is the mean of the exit assessment readings for each domain. In the domain of *Physical health*, the average reading at service exit for Kōkihi ngā Rito tamariki is 4.42.

Average change

This is the mean difference between entry and exit readings. For tamariki supported by Kōkihi ngā Rito, the average difference between entry and exit readings in *Physical health* is 0.79.

Standard deviation

This measures how much individual changes typically deviate from the mean change. With a possible range of change values from -5 to 5, a standard deviation of 1.21 (as seen in *Physical health*) indicates relatively low dispersion, suggesting that most children's change journeys cluster fairly closely together.

SEM (standard error of the mean)

The standard error of the mean (SEM) is a value that measures the precision of the sample mean as an estimate of the population mean. The *Physical health* SEM of 0.14 indicates that the sample mean change of 0.79 likely falls within ± 0.14 of the true population mean.

t (t-value)

The t-statistic measures the size of the difference between entry and exit readings relative to the variation in the data. A positive value indicates that exit readings were higher than entry readings on average. For example, *Physical health* shows a t-value of 5.51.

p (p-value)

The p-value indicates the probability of observing these changes if there was no real difference between the entry and exit readings. For example, the p-value of <0.0001 for *Physical health* means there is less than a 0.01% chance of observing these differences by chance, if there is no difference between entry and exit. This is well below the conventional 0.05 threshold for statistical significance.

95% confidence interval

The 95% confidence interval shows the range within which we can be 95% confident that the true population mean change lies. The interval (0.50 to 1.07) for *Physical health* indicates we can be 95% confident that the true average change falls between 0.50 and 1.07 units.

Cohen's d

This is a standardised measure of effect size. A Cohen's d value indicates how many standard deviations of change generally occurred between entry and exit readings. For example, in *Physical health*, a Cohen's d of 0.77 means that on average, each child's exit reading was 0.77 standard deviations higher than their entry reading.

Effect size

Effect size categories help us understand the practical significance of the observed changes. According to Cohen[40], "a medium effect of 0.5 is visible to the naked eye of a careful observer." In *Physical health*, the effect size of 0.77 represents a medium-to-large effect, which approaches the threshold of 0.8 for a large effect.

Outcomes Star™ My Star™ pre-post analysis:

Tamariki supported by other programmes

Statistical methods

After describing the wellbeing changes observed in tamariki between the start and end of the Kōkihi ngā Rito programme, we conducted the exact same set of analyses on the tamariki who were not participating in the programme. This helped establish a clear baseline of impact for both groups before continuing to assess whether the impact of the Kōkihi ngā Rito programme is greater.

Again, a series of paired t-tests were conducted across all eight domains of the My Star™, comparing the readings for the same individuals (or matched pairs) before and after the service intervention (or treatment). As with the Kōkihi ngā Rito cohort, initial examination of the data confirmed the absence of outliers and a reasonable approximation of normality.

Interpretation of findings

As shown in Table 2, the results suggest that participation in programmes other than Kōkihi ngā Rito also had a significant and medium-to-large effect on the wellbeing of tamariki across seven of the eight domains measured. Specifically, the participants saw a medium effect in the domains of *Feelings and behaviour*, and *Confidence and self-esteem*; and a large effect in the domains of *Physical health, Where you live, Being safe, Friends, Education and learning*, and in the overall average across all domains. The smallest observed effect size with a statistically significant p-value was 0.47, or a medium effect, on *Confidence and self-esteem*. All effect sizes were positive, and p-values were significant for all domains except *Relationships*.

Limitations

As with the Kōkihi ngā Rito cohort, the limitations of the paired t-test itself include the inability to control for confounding variables and a focus on just the mean difference, which prevents any inference of causality.

Conclusions

The paired t-test analyses suggests that support provided by NCIWR contracts outside of the Kōkihi ngā Rito programme is also associated with medium to large and statistically significant increases in wellbeing across seven different domains. The general agreement across My Star™ domains coupled with statistical significance and effect size combine to provide solid evidence for the impact of non-Kōkihi ngā Rito service on the wellbeing of this specific cohort of tamariki.

Although comparison of My Star™ data between the Kōkihi ngā Rito cohort with the non-Kōkihi ngā Rito cohort is influenced by the limitations discussed in the tamariki cohort identification section, we nonetheless note larger effect sizes across most domains for the Kōkihi ngā Rito tamariki, as shown in Table 3.

Table 2. Paired t-test results for non-Kōkihi ngā Rito tamariki (n=43)

My Star™ Domain	Average initial	Average final	Average change	Standard deviation	SEM	t	p	95% confidence interval	Cohen's d	Effect size
Physical health	3.56	4.40	0.84	1.04	0.16	5.25	<0.0001*	(0.52 - 1.16)	0.90	Large
Where you live	3.26	4.33	1.07	1.42	0.22	4.94	<0.0001*	(0.63 - 1.51)	0.87	Large
Being safe	3.30	4.30	1.00	1.13	0.17	5.78	<0.0001*	(0.65 - 1.35)	0.95	Large
Relationships	3.79	4.16	0.37	1.35	0.21	1.81	0.0769	(-0.04 - 0.79)	0.35	Small
Feelings and behaviour	3.37	4.00	0.63	1.18	0.18	3.50	0.0011*	(0.27 - 0.99)	0.62	Medium
Friends	4.07	4.70	0.63	0.82	0.12	5.04	<0.0001*	(0.38 - 0.88)	0.85	Large
Confidence and self-esteem	3.86	4.33	0.47	1.18	0.18	2.58	0.0135*	(0.10 - 0.83)	0.47	Small-to-medium
Education and learning	3.74	4.44	0.70	1.10	0.17	4.15	0.0002*	(0.36 - 1.04)	0.71	Medium

Table 3. Comparison of effect sizes between tamariki supported by Kokihi ngā Rito (n=71) and other programmes (n=43)

	Kōkihi ngā Rito		Other programmes	
My Star™ Domain	Cohen's d	Effect size	Cohen's d	Effect size
Physical health	0.77	Medium-to-large	0.90	Large
Where you live	1.28	Large-to-very large	0.87	Large
Being safe	1.21	Large-to-very large	0.95	Large
Relationships	0.94	Large	0.35†	Small†
Feelings and behaviour	1.44	Very large	0.62	Medium
Friends	0.75	Medium-to-large	0.85	Large
Confidence and self-esteem	1.10	Large	0.47	Small-to-medium
Education and learning	0.46	Small-to-medium	0.71	Medium

†Relationships is not significant for the Relationship domain, in the non-Kōkihi ngā Rito cohort.

Outcomes Star™ My Star™ comparison analyses:

Propensity Score Matching and Difference-in-Differences

Statistical methods

After establishing that wellbeing changes are observed in both tamariki cohorts (those supported by Kōkihi ngā Rito and those supported by other services), we moved on to address the first key research question: are the wellbeing changes greater for tamariki in the Kōkihi ngā Rito programme?

To explore this question, we conducted both Propensity Score Matching (PSM) analyses and Difference-in-Differences (DiD) analyses. These two approaches target slightly different goals, rely on different assumptions, and have distinct limitations. By using both methodologies and assessing the concordance and discrepancies between results, we aimed to better understand and account for any weaknesses in the dataset[41].

PSM is a method used to compare two groups (typically a treatment and a control group) by matching individuals who are similar based on observable characteristics (e.g., age, ethnicity, deprivation index). It aims to reduce bias in observational studies by ensuring that the groups being compared are as similar as possible, and it works by matching individuals from the treatment group with individuals from the control group who have similar propensity scores (probabilities of receiving the treatment based on their characteristics)[42,43]. PSM is useful when you want to control for differences between groups that could affect the outcome but don't have random assignment, such as the case here. Tamariki were not randomly assigned to either the Kōkihi ngā Rito or non-Kōkihi ngā Rito cohort, and PSM attempts to ensure that we compare the most similar tamariki between the groups.

DiD, on the other hand, is a method that compares the change in outcomes over time between a treatment group and a control group, assuming that, in the absence of treatment, both groups would have followed the same trend over time. DiD looks at the difference in outcomes before and after a treatment (or intervention) in both groups and subtracts the pre-treatment difference from the post-treatment difference to isolate the treatment effect[44,45]. It is particularly useful (and therefore widely used) for evaluating the impact of a policy or intervention where data is available both before and after the intervention for both treated and untreated groups, as is also the case here. Ideally, the two approaches can be combined, with PSM balancing cohorts and DiD determining the different treatment effects between them.

In addition to combining analytical approaches, we also performed the same calculations with both the "Inclusion group" and "Exclusion group" of non-Kōkihi ngā Rito tamariki. Because there are legitimate reasons to argue for or against including certain tamariki in the control, we sought to understand whether results were sensitive to those choices on the part of the data scientist.

Interpretation of findings

Across all PSM and DiD analyses, the *Feelings and behaviour* domain returned consistently statistically significant treatment effects. Similarly, the *Relationships* domain was significant in all tests except for the exclusion DiD, where its p-value crept up to 0.06 – not statistically significant at the official level of 5%, but borderline.

The DiD analyses using the inclusion cohort suggest that the *Confidence and self-esteem* domain is also statistically significant. This domain is close to significant in the exclusion PSM ($p = 0.081$) but not in the inclusion PSM ($p = 0.20$); interestingly, the trend is the opposite with DiD analyses, with the domain significant in the inclusion DiD ($p=0.04$) but borderline in the exclusion DiD ($p=0.05$).

Both PSM and DiD analyses saw the *Where you live* domain vault to statistically significant when using the exclusion cohort. Given that the treatment effect in this domain is highly sensitive to the inclusion of 12 specific tamariki, we recommend a cautious interpretation of this result.

Housing instability is a common experience entwined with family violence[46,47,48,49,50], but the nuance of that experience may be diverse across tamariki. Particularly with small sample sizes, it is possible that some of the 31 tamariki left in the exclusion cohort were experiencing particularly difficult housing situations that improved during service, and that this signature is simply washed out when we add back the 12. Other NCIWR services, particularly those supporting the mothers of Kōkihi ngā Rito tamariki, help wāhine secure new, safe housing. While that is not the primary work of the Kōkihi ngā Rito programme, we are certain to see a downstream impact on tamariki wellbeing. In the context of the housing crisis in Aotearoa and its impacts in the sphere of family violence, which operate as obvious confounders, we would advise caution in making any causal inference within this domain. We would not interpret these results as indicative of a causal relationship directly between the Kōkihi ngā Rito programme and the *Where you live* domain of tamariki wellbeing but suggest that this is an area that warrants further investigation.

Conversely, the Kōkihi ngā Rito programme is explicitly designed to support tamariki with their feelings, relationships, and self-esteem, areas that correspond to the three other My Star™ domains with strong treatment effects. The programme logic, coupled with the statistically significant results in these domains across multiple methodologies and multiple control cohorts, provides robust evidence that participation in the Kōkihi ngā Rito programme results in greater wellbeing gains than other NCIWR services for children.

Limitations

PSM is designed to match individuals based solely on the variables included in the model and cannot control for unobserved confounders. The results in the Where you live domain suggest that at least one area has confounders present, but we cannot know whether other unobserved variables may have biased the treatment effect. The use of the nearest neighbour matching algorithm was selected to guard against another PSM limitation, imperfect matching resulting in loss of sample size, but this approach does mean that certain tamariki were used as matches more than once and any confounder present in these children's experiences could have been magnified.

The DiD analyses are based on a key assumption of parallel trends – that is, in the absence of any treatment, the two groups would follow the same (or parallel) trend. While we have no evidence to dispute this assumption, it is equally difficult to check for violations. We know that tamariki were not randomly assigned to treatment groups, so it is possible that pre-treatment differences in the groups could account for some of the observed treatment effect.

Conclusions

The results of multiple PSM and DiD analyses indicate that participation in the Kōkihi ngā Rito programme had a significant impact on the wellbeing of tamariki across domains of *Feelings and behaviour*, *Relationships*, and *Confidence and self-esteem*. The treatment effects in each of these three domains, as measured by the differential change between the Kōkihi ngā Rito tamariki and two different control cohorts, suggest that the intervention was effective. With larger sampling, further analyses could explore potential heterogeneity in effects across subgroups and examine long-term outcomes or explore potential unmeasured confounding factors, particularly in the areas of wellbeing directly impacted by housing insecurity or insufficiency.

Table 4. Propenstiy Score Matching results

a) Kōkihi ngā Rito (n=71) vs "Inclusion" cohort (n=43)

My Star™ Domain	ATT	Std Error	Z	p	95% confidence interval
Physical health	-0.43	0.33	-1.29	0.1960	(-1.08 - 0.22)
Where you live	0.17	0.38	0.44	0.6590	(-0.57 - 0.91)
Being safe	0.09	0.27	0.32	0.7500	(-0.45 - 0.62)
Relationships	**0.64**	**0.29**	**2.22**	**0.0260***	**(0.08 - 1.20)**
Feelings and behaviour	**0.61**	**0.26**	**2.32**	**0.0200***	**(0.09 - 1.12)**
Friends	0.23	0.23	1.00	0.3170	(-0.22 - 0.69)
Confidence and self-esteem	0.35	0.27	1.26	0.2070	(-0.19 - 0.88)
Education and learning	-0.18	0.26	-0.67	0.5050	(-0.69 - 0.34)

Bold indicates statistical significance at p=0.05

b) Kōkihi ngā Rito (n=71) vs "Exclusion" cohort (n=31)

My Star™ Domain	ATT	Std Error	Z	p	95% confidence interval
Physical health	0.36	0.26	1.37	0.1720	(-0.16 - 0.87)
Where you live	**0.90**	**0.25**	**3.67**	**0.0000***	**(0.42 - 1.38)**
Being safe	0.43	0.28	1.51	0.1310	(-0.13 - 0.98)
Relationships	**0.59**	**0.30**	**1.99**	**0.0470***	**(0.01 - 1.18)**
Feelings and behaviour	**0.87**	**0.24**	**3.61**	**0.0000***	**(0.40 - 1.35)**
Friends	0.48	0.28	1.69	0.0920	(-0.08 - 1.03)
Confidence and self-esteem	0.48	0.27	1.75	0.0810	(-0.06 - 1.02)
Education and learning	-0.08	0.31	-0.25	0.8060	(-0.67 - 0.52)

Bold indicates statistical significance at p=0.05

A guide to interpretation:

ATT

This is the Average Treatment Effect on the Treated (ATT), which measures the average difference in outcomes between the treatment and matched control group. The ATT of *Physical health* is -0.43. The negative sign indicates a negative effect of treatment; in other words, that participants in Kōkihi ngā Rito saw, on average, a 0.43 smaller change in their *Physical health* readings.

Std Error

This is the standard error of the ATT, which quantifies the precision of the estimate – smaller standard errors indicate more precise estimates. For *Physical health*, the standard error is 0.33, which is nearly as large in magnitude as the ATT itself, indicating a relatively high level of uncertainty associated with the measure of ATT.

Z (z-statistic)

This is the calculated z-statistic, which is -1.29 for the *Physical health* domain. A z-statistic is used to test whether the ATT is statistically different from zero and indicates that the ATT is approximately 1.29 standard deviations below the mean. In other words, the estimated effect is not very far from zero.

p (p-value)

The p-value indicates the probability that the observed treatment effect is due to chance; in other words, if the treatment effect is zero, how likely would we be to see these results? The p-value of 0.196 for *Physical health* means there is roughly a 20% chance of observing these test results if there's no actual treatment effect. Typically, a p-value of less than 0.05 is considered statistically significant, so this result suggests that there is no statistically significant difference in the treatment effect of Kōkihi ngā Rito service compared to other NCIWR programmes on the *Physical health* domain of wellbeing.

95% confidence interval

This is the 95% confidence interval for the ATT, which is an estimate of the range of values within which the true treatment effect is likely to lie. The interval of -1.08–0.22 for *Physical health* indicates that we can say with 95% confidence that the true treatment effect sits between -1.08 and 0.22. This interval includes zero, which represents no treatment effect. In other words, we have no evidence to support any greater effect of Kōkihi ngā Rito on the *Physical health* domain of tamariki wellbeing compared to other NCIWR services.

Table 5. Difference-in-Differences results

a) Kōkihi ngā Rito (n=71) vs "Inclusion" cohort (n=43)

My Star™ Domain	Treatment effect	Std Error	t	p	95% confidence interval
Physical health	-0.05	0.27	-0.18	0.86	(-0.58 - 0.48)
Where you live	0.38	0.32	1.19	0.24	(-0.25 - 1.01)
Being safe	0.30	0.29	1.02	0.31	(-0.28 - 0.87)
Relationships	**0.66**	**0.30**	**2.22**	**0.03***	**(0.07 - 1.24)**
Feelings and behaviour	**0.61**	**0.25**	**2.43**	**0.02***	**(0.12 - 1.11)**
Friends	0.17	0.26	0.67	0.50	(-0.34 - 0.69)
Confidence and self-esteem	**0.53**	**0.26**	**2.09**	**0.04***	**(0.03 - 1.04)**
Education and learning	-0.19	0.29	-0.66	0.51	(-0.76 - 0.38)

Bold indicates statistical significance at p=0.05

b) Kōkihi ngā Rito (n=71) vs "Exclusion" cohort (n=31)

My Star™ Domain	Treatment effect	Std Error	t	p	95% confidence interval
Physical health	0.12	0.31	0.40	0.69	(-0.48 - 0.72)
Where you live	**0.75**	**0.34**	**2.18**	**0.03***	**(0.07 - 1.43)**
Being safe	0.60	0.31	1.92	0.06	(-0.01 – 1.21)
Relationships	0.63	0.33	1.90	0.06	(-0.03 - 1.28)
Feelings and behaviour	**0.57**	**0.26**	**2.20**	**0.03***	**(0.06 - 1.09)**
Friends	0.34	0.30	1.13	0.26	(-0.25 - 0.92)
Confidence and self-esteem	0.53	0.27	1.95	0.05	(-0.01 - 1.07)
Education and learning	-0.23	0.33	-0.69	0.49	(-0.87 - 0.42)

Bold indicates statistical significance at p=0.05

A guide to interpretation:

Treatment effect

The treatment effect represents the causal impact of an intervention. For *Physical health*, the treatment effect is -0.05. Like ATT, the sign indicates directionality of effect, and the absolute value represents its magnitude. This indicates that participants in Kōkihi ngā Rito saw, on average, a 0.05 smaller change in their *Physical health* readings.

Std Error

This is the standard error of the treatment effect, which quantifies the precision of the estimate – smaller standard errors indicate more precise estimates. For *Physical health*, the standard error is 0.27, which is nearly five times greater in magnitude than the treatment effect itself; this indicates a very high level of variability or uncertainty associated with the measure of treatment effect in this domain.

t (t-statistic)

This is the t-statistic, which quantifies the difference between the estimated treatment effect and the null hypothesis (in this case, that the treatment effect is zero). The value of -0.18 for the *Physical health* domain indicates that the estimated treatment effect is just 0.18 standard deviations below zero, the null hypothesis.

p (p-value)

Exactly as in PSM, the p-value indicates the probability that the observed treatment effect is due to chance; in other words, if the treatment effect is zero, how likely would we be to see these results? The p-value of 0.86 for *Physical health* means there is more than an 80% chance of observing these test results if there is no actual treatment effect, which suggests that there is no treatment effect of Kōkihi ngā Rito on the *Physical health* domain of tamariki wellbeing.

95% confidence interval

This is the 95% confidence interval for the treatment, which provides an estimate of the range of values within which the true treatment effect is likely to lie. The interval of -0.58 – 0.48 for *Physical health* indicates that we can say with 95% confidence that the true treatment effect sits within this range. As the range is essentially centred on zero, this is further indication that we cannot reject the null hypothesis; in other words, we still have no evidence to support any greater effect of the Kōkihi ngā Rito programme on the *Physical health* domain of tamariki wellbeing compared to other NCIWR services.

Outcomes Star™ Empowerment Star™ analyses

Statistical methods

The aim of these analyses was to understand whether providing support for children has an influence on a mother's wellbeing. A series of Kruskal-Wallis tests were conducted to compare the distribution of change (the difference between initial and final collections of the Empowerment Star™) across nine different domains of wellbeing, between three independent cohorts of wāhine: mothers with tamariki supported by Kōkihi ngā Rito, mothers with tamariki supported by other NCIWR programmes, and mothers with tamariki never directly supported by any NCIWR programme.

A non-parametric Kruskal-Wallis test was chosen in favour of the parametric ANOVA due to the heteroscedasticity of the data. Parametric tests like ANOVA and t-test are generally quite robust to violations of normality, but sensitive to unequal variances between groups, especially when sample sizes are unequal, as is the case here [51,52,53].

Interpretation of findings

The statistical significance of the *Children* domain of wellbeing for cohorts of wāhine whose tamariki received different levels of support from NCIWR was not surprising; however, the specific areas where that difference occurs indicate that this is not necessarily an effect of the Kōkihi ngā Rito programme.

As shown in Table 7, the largest difference in wellbeing gains was observed when comparing the cohort of wāhine whose tamariki received non-Kōkihi ngā Rito support from NCIWR with those wāhine whose tamariki received no directed service provision from NCIWR. Providing any support to tamariki corresponds to a greater increase in their mother's wellbeing with regards to her children.

The absence of statistically significant differences in the *Children* domain between mothers with tamariki in Kōkihi ngā Rito and those whose children received support elsewhere may stem from several confounding factors. Children referred to Kōkihi ngā Rito likely represent cases with heightened vulnerability and risk, which inevitably affects maternal wellbeing in this domain. When children face ongoing threats from a perpetrator (typically their father), this danger often persists regardless of support services provided. The complexities of family court proceedings compound these challenges. When abusers maintain legal access to child victims of family violence, this significantly impacts maternal wellbeing. To fully assess this domain, a longer observation period would be valuable; while Kōkihi ngā Rito offers child-centred support, a mother's concerns about her children typically persist until comprehensive legal and physical protections are established. If children supported through alternative programmes face lower initial vulnerability levels, it's reasonable to suggest that external risks may have already been mitigated in those cases, thus enabling child-focused interventions to more effectively improve maternal wellbeing.

The *Work and learning* domain returns a similar result to the *Children* domain; in fact, the result is even more statistically significant. The interaction of support for tamariki and a mother's wellbeing in terms of work and education is not immediately obvious, but Kaiārahi Tamariki for the Kōkihi ngā Rito programme often collect tamariki from school. For example, one describes practicing safety planning with tamariki as "sometime when I pick the kids up from school, I don't let them in the van till they can recite parts of their safety plans," and a tamariki from the programme described their Kaiārahi Tamariki as "always picking us up after school, she takes us to our after-school things"[29]. It is possible that this additional support may help wāhine to maintain or regain work or education. As with the *Children* domain, however, the large and significant difference is observed when comparing wāhine whose children received no targeted support with those whose children were served by other NCIWR programmes. We recommend exploring with Advocates and Kaiārahi tamariki what other activities are part of the services that support children, and whether any other elements of those programmes could have a logical connection to a mother's employment or education journey. Although the data indicate a strong relationship, the voices of lived experience are likely to prove more valuable in illuminating the real nature of the connection observed here.

Table 6. Kruskal-Wallis χ^2 results, by Empowerment Star™ domain

Empowerment Star™ Domain	χ2	p
Safety	0.41	0.815
Accommodation	0.10	0.952
Support networks	1.32	0.518
Legal issues	2.52	0.283
Health and wellbeing	4.10	0.129
Money	4.38	0.112
Children	**9.58**	**0.008***
Work and learning	**10.93**	**0.004***
Empowerment and self-esteem	3.24	0.198

A guide to interpretation:

χ^2 (Chi-squared)

This is the test statistic for the Kruskal-Wallis test. The larger the value, the more evidence there is for differences between the groups, while a lower value suggests that the differences between the groups are less pronounced. The χ^2 value of 0.41 for *Safety* indicates very little observable difference between wāhine cohorts regarding changes in this wellbeing domain.

p (p-value)

The p-value represents the probability of observing the data, or something more extreme, if the null hypothesis is true. The null hypothesis for the Kruskal-Wallis test is that all groups are equal, so the p-value of 0.815 for *Safety* suggests that there is no difference between any of the three wāhine cohorts in this domain, and so no further tests are warranted. In the domain of *Children*, however, the p-value on 0.008 indicates statistically significant differences between the three cohorts, and so additional Dunn-Bonferroni tests should be conducted to identify the specific nature of those differences. These results are included in Table 7 on the following page.

Limitations

By design, the Dunn-Bonferroni test is sensitive to sample size. The more comparisons you make, the more likely you are to return false positives (Type I errors). Consequently, the Bonferroni correction intentionally reduces the significance level as the number of comparisons increases. For small sample sizes, this can drastically reduce the power of the test, increasing the risk of Type II errors (failing to detect real effects). With larger sample sizes, the impact is less severe because the increased precision helps mitigate the stricter significance threshold.

In our wāhine cohorts, we have 2,682 wāhine whose tamariki have never had a referral to NCIWR and 398 wāhine whose children have had a referral into other services, but just 54 wāhine with a Kōkihi ngā Rito tamariki as well as the two Empowerment Star™ readings necessary for analysis. Therefore, we may not have a large enough cohort of Kōkohi ngā Rito mothers to achieve statistical power, and the profound influence of providing any support at all to a mother's tamariki may have washed out any signal of difference in impact according to the programme that delivers that tamariki support.

Conclusion

The results of the Kruskal-Wallis and post-hoc Dunn-Bonferroni tests indicate that mothers whose tamariki are also supported by NCIWR experience greater increases in their own wellbeing within the domains of both *Children* and *Work and learning*. Although the effect is observed regardless of which programme supports the tamariki, our ability to differentiate between Kōkihi ngā Rito and other programmes may be a factor of the sampling. In the future, larger sample sizes of mothers with tamariki in Kōkihi ngā Rito may help to clarify any difference in impact. Further analyses could also incorporate surveys, focus groups, and mixed methods to investigate potential mechanisms for the apparent relationship between tamariki support and a mother's wellbeing in the domain of *Work and learning*.

Table 7. Dunn-Bonferroni results, by Empowerment Star™ domain

a) *Children* domain

Cohort comparison	Test statistic	Std error	Std test statistic	p-value	Adj p-value
KNR mum <> No children support mum	-45.63	122.64	-0.37	0.71	1
KNR mum <> Other service mum	-191.83	129.40	-1.48	0.138	0.415
Other service mum <> No children support mum	**146.20**	**47.93**	**3.05**	**0.002**	**0.007***

b) *Work and learning* domain

Cohort comparison	Test statistic	Std error	Std test statistic	p-value	Adj p-value
KNR mum <> No children support mum	-154.35	121.95	-1.27	0.206	0.613
KNR mum <> Other service mum	-296.69	128.67	-2.31	0.021	0.063
Other service mum <> No children support mum	**142.43**	**47.66**	**2.99**	**0.003**	**0.008***

A guide to interpretation:

Test statistic

This is the test statistic for Dunn's test. The sign of the result indicates directionality, and the absolute value represents the magnitude of difference. The result of -45.63 for the pairwise comparison between *KNR mum <> No children support mum* indicates that wellbeing change in the domain of *Children* is slightly smaller for mothers with tamariki in Kōkihi ngā Rito when compared to mothers whose children have received no support.

Std error

This is the standard error of the test statistic, which indicates the precision of the result – smaller standard errors indicate more precise estimates. For the comparison of *KNR mum <> No children support mum*, the standard error is 122.64, over twice as large in magnitude as the test statistic itself. This indicates a reasonable amount of variability in the measurement of that statistic.

Std test statistic

This is the standardised test statistic, which is simply the original statistic adjusted to correct for differences between group ranks due to variability in the data. Mathematically, it is the difference between the mean ranks of the two groups divided by the standard error, and functionally, it is essentially a z-statistic. It represents how far, in units of standard deviations, a data point sits from the expected mean under the null hypothesis. The value of -0.37 for the comparison between *KNR mum <> No children support mum* is small, indicating that this result sits just 0.37 standard deviations from zero, our null hypothesis.

p-value

The p-value represents the probability that the observed treatment effect is due to chance. In other words, if the treatment effect is zero, how likely would we be to see these results? The p-value of 0.71 means there is more than a 70% chance of observing these test results if there is no difference between any of the three wāhine cohorts.

Adjusted p-value

This is the adjusted p-value, which modifies the calculated p-value to account for the multiple comparisons being performed. This reduces the risk of false positives, or Type I errors, by making the significance threshold more stringent. Since the original p-value of 0.71 was already not significant, it is unsurprising that it becomes even less significant after correction, returning 1. This indicates that there is absolutely no evidence to support any difference in wellbeing change in the domain of *Children* when comparing mothers of Kōkihi ngā Rito tamariki with mothers whose children have never been supported by a NCIWR programme.

Service delivery analyses

Statistical methods

The Kōkihi ngā Rito programme includes unique service delivery elements that make it distinct from other NCIWR services that may support tamariki, and some of these differences may act as confounders when exploring case notes across tamariki cohorts. Therefore, these analyses aimed to understand the nature, magnitude, and distribution of this variability. For each referral, we compared referral duration, count of support activities, total support hours, count of case notes, and total case note characters between the three tamariki cohorts.

Initial analyses returned a strong and significant difference in referral duration, so support activity and case note measures were adjusted to "per referral" measures to correct for this—total support hours per referral, and total case note characters per referral.

The number of activities was strongly correlated with the number of support hours, but the latter was selected as a better proxy for the quantity of support delivered, because activities, such as attempts to contact, are often recorded. While these brief interactions are undoubtedly necessary to provide care, they inflate the activity count and do not necessarily represent direct tamariki support. The actual hours spent with tamariki was determined to be a better representation.

Because of the robust internal documentation procedures at NCIWR, the number of activities and the number of case notes were nearly identical – Advocates and Kaiārahi tamariki wrote a note for almost every single recorded interaction or activity with tamariki. Therefore, we discarded the count of case notes as a separate measure and focused instead on the contents of those case notes, with their comprehensiveness approximated by the number of characters in each.

Once measures were reduced to just three, a series of Kruskal-Wallis tests were conducted to compare their distribution between three independent cohorts of tamariki: participants in the Kōkihi ngā Rito programme, participants in other NCIWR programmes delivered at a refuge that was also delivering Kōkihi ngā Rito, and tamariki supported by a refuge that does not offer Kōkihi ngā Rito. As with the wāhine Empowerment Star™ data, the unequal variances observed in this dataset prompted the selection of a non-parametric test.

These analyses were also conducted within two overlapping groups to explore whether differing risk profiles within the cohorts could influence observed differences. The initial analyses included all tamariki of appropriate ages, and a second set of tests restricted all cohorts down to just those tamariki whose mothers had reported all six child-related risk items on her risk assessment. This was intended to provide a degree of control for the possibility that Kōkihi ngā Rito tamariki are simply the highest need or highest risk tamariki, which would be expected to impact service delivery practices.

Table 8. Mean and confidence interval for service delivery measures, by tamariki cohort

Measure	Cohort	Mean	95% confidence interval
Days in referral	KNR	186.90	164.08 - 209.72
	nonKNR at KNR refuge	102.71	99.8 - 105.61
	nonKNR refuge	111.63	109.39 - 113.87
Average daily support time	KNR	46.8	37.2 - 56.4
	nonKNR at KNR refuge	9.0	8.4 - 9.0
	nonKNR refuge	6.6	6.0 - 7.2
Average case note length	KNR	5,616.37	4,409.49 - 6.823.25
	nonKNR at KNR refuge	566.08	534.28 - 597.88
	nonKNR refuge	411.63	394.59 - 428.66

Table 9. Kruskal-Wallis and Dun-Bonferroni results, by service delivery measure

a) Days per referral, for all tamariki (n=26,362)

χ2	df	p-value	η2
217.52	**2**	**<.001***	**0.009**

Cohort comparison	Test statistic	Std error	Std test statistic	p-value	Adj p-value
KNR <> nonKNR at KNR refuge	**5,411.07**	**575.55**	**9.40**	**<.001**	**<.001***
KNR <> nonKNR refuge	**6,519.92**	**568.66**	**11.45**	**<.001**	**<.001***
nonKNR at KNR refuge <> nonKNR refuge	**1,099.85**	**113.08**	**9.73**	**<.001**	**<.001***

Bold indicates statistical significance at p=0.05

b) Support hours per day, for all tamariki (n=23,632)

χ2	df	p-value	η2
447.91	**2**	**<.001***	**0.018**

Cohort comparison	Test statistic	Std error	Std test statistic	p-value	Adj p-value
KNR <> nonKNR at KNR refuge	**6,398.18**	**569.03**	**11.24**	**<.001**	**<.001***
KNR <> nonKNR refuge	**8,177.64**	**562.21**	**14.55**	**<.001**	**<.001***
nonKNR at KNR refuge <> nonKNR refuge	**1,779.45**	**111.80**	**15.92**	**<.001**	**<.001***

Bold indicates statistical significance at p=0.05

c) Characters per case note, for all tamariki (n=23,632)

χ2	df	p-value	η2
1,406.34	**2**	**<.001***	**0.059**

Cohort comparison	Test statistic	Std error	Std test statistic	p-value	Adj p-value
KNR <> nonKNR at KNR refuge	**8,942.19**	**474.23**	**18.86**	**<.001**	**<.001***
KNR <> nonKNR refuge	**11,645.19**	**468.56**	**24.85**	**<.001**	**<.001***
nonKNR at KNR refuge <> nonKNR refuge	**2,703.00**	**93.18**	**29.01**	**<.001**	**<.001***

Bold indicates statistical significance at p=0.05

A guide to interpretation:

χ2

Sometimes seen written as Chi-squared or Chi^2, this is the test statistic for the Kruskal-Wallis test. The larger the value, the more evidence there is for differences between the groups, while a lower value usually suggests that the differences between the groups are not as pronounced. The χ^2 value of 217.52 for *Days per referral* indicates a sizeable amount of difference between tamariki cohorts for this measure.

df

The degrees of freedom (df) is used in significance calculations. It represents the number of independent variables that are free to vary in order to calculate a particular statistic. It is equal to the number of groups minus 1, because once you know all of the values except the last, that 1 cannot vary. With three cohorts or groups, the degrees of freedom is 3 - 1 = 2.

p-value

The p-value represents the probability of observing the data, or something more extreme, if the null hypothesis is true. The null hypothesis for the Kruskal-Wallis test is that all groups are equal, so the p-value of <0.001 for *Days in this referral* provides strong evidence that we should reject that null hypothesis; in other words, there are some significant differences between these cohort groups.

η²

Pronounced "eta-squared," this is a measure of effect size, which provides an estimate of the proportion of variance explained by the group differences. Like most effect size measures, values typically range from 0 to 1, with higher values indicating larger effects. The effect size of 0.009 is small, while a value near 0.06 would be medium, and 0.14 is generally interpreted as a large effect size.

Test statistic

This is the test statistic, calculated for each of the pairwise comparisons between tamariki cohorts. The sign of the result

Table 9. Kruskal-Wallis and Dun-Bonferroni results, by service delivery measure *(continued)*

indicates directionality, and the absolute value represents the magnitude of difference. The result of 5,441.07 for the pairwise comparison between KNR <> nonKNR at KNR refuge for the *Days in referral* measure indicates a sizeable difference between these groups and tells us that referral duration is longer on average in the Kōkihi ngā Rito group.

Std error

This is the standard error of the test statistic, which indicates the precision of the result – smaller standard errors indicate more precise estimates. For the comparison of KNR <> nonKNR at KNR refuge for the *Days in referral* measure, the standard error is 575.55, roughly 10% as large in magnitude as the test statistic itself. This indicates relatively little variability in the measurement of that statistic.

Std test statistic

This is the standardised test statistic, which is simply the original statistic adjusted to correct for differences between group ranks due to variability in the data. The value of 9.40 for the comparison between KNR <> nonKNR at KNR refuge for the *Days in referral* measure is large for a z-statistic, indicating that this result sits more than nine standard deviations from zero.

p-value

The p-value in the lower section of the table represents the probability that the observed treatment effect is due to chance; in other words, if these two cohorts of tamariki have the same average referral duration, how likely would we be to see these results? The p-value of <0.001 means that we would be very unlikely to observe this data in that situation and provides evidence that we should reject the null hypothesis.

Adjusted p-value

This is the adjusted p-value, which modifies the calculated p-value to reduce the risk of false positives (Type I errors). After correction, the p-value remains <0.001, so we reject the null hypothesis that referral durations are the same between Kōkihi ngā Rito tamariki and other programmes at the same refuge.

d) Days per referral, for highest risk profile tamariki (n=1,733)

		χ2	df	p-value	η2
		42.71	**2**	**<.001***	**0.023**
Cohort comparison	Test statistic	Std error	Std test statistic	p-value	Adj p-value
KNR <> nonKNR at KNR refuge	**435.10**	**73.07**	**5.95**	**<.001**	**<.001***
KNR <> nonKNR refuge	**465.44**	**71.34**	**6.52**	**<.001**	**<.001***
nonKNR at KNR refuge <> nonKNR refuge	30.34	26.35	1.15	0.25	0.749

Bold indicates statistical significance at p=0.05

e) Support hours per day, highest risk profile tamariki (n=1,733)

		χ2	df	p-value	η2
		106.97	**2**	**<.001***	**0.061**
Cohort comparison	Test statistic	Std error	Std test statistic	p-value	Adj p-value
KNR <> nonKNR at KNR refuge	**420.23**	**72.52**	**5.79**	**<.001**	**<.001***
KNR <> nonKNR refuge	**600.30**	**70.81**	**8.48**	**<.001**	**<.001***
nonKNR at KNR refuge <> nonKNR refuge	**180.07**	**26.15**	**6.89**	**<.001**	**<.001***

Bold indicates statistical significance at p=0.05

f) Characters per case note, highest risk profile tamariki (n=1,733)

		χ2	df	p-value	η2
		326.13	**2**	**<.001***	**0.19**
Cohort comparison	Test statistic	Std error	Std test statistic	p-value	Adj p-value
KNR <> nonKNR at KNR refuge	**300.25**	**42.56**	**7.05**	**<.001**	**<.001***
KNR <> nonKNR refuge	**437.54**	**41.40**	**10.57**	**<.001**	**<.001***
nonKNR at KNR refuge <> nonKNR refuge	**137.29**	**19.43**	**7.07**	**<.001**	**<.001***

Bold indicates statistical significance at p=0.05

Interpretation of findings

The results of Kruskal-Wallis tests show a statistically significant difference in the days per referral across the three full tamariki cohorts ($\chi 2 = 42.71$, $p = <0.001$). Post-hoc pairwise comparisons revealed that tamariki supported by Kōkihi ngā Rito had significantly longer referrals compared to both non-Kōkihi ngā Rito cohorts, those at a KNR refuge ($p = <0.001$) and those at a different refuge ($p=<0.001$). When all tamariki were included, tests also indicated a difference in referral duration between the two non-Kōkihi ngā Rito tamariki cohorts, despite nearly identical mean days in service. This difference stems from the much larger variance within refuges that do not deliver Kōkihi ngā Rito but have a long tail of outliers in the form of long-duration referrals. When we reduce sampling down to just those tamariki identified as "higher risk " (with all six child-related risk items reported by their mother), the significance of the difference between these two cohorts is lost. Within the high-risk comparison groups, there was no significant difference between the two non-Kōkihi ngā Rito cohorts, suggesting that the difference in days in referral is likely related to the Kōkihi ngā Rito programme itself rather than refuge-specific factors.

The results for support hours per day show a significant difference across all tamariki cohorts, regardless of whether we consider all tamariki or just those with higher risk profiles (all tamariki: $\chi 2 = 447.91$, $p = <0.001$; highest risk: all tamariki: $\chi 2 = 106.94$, $p = <0.001$) with Kōkihi ngā Rito having the highest support hours per day compared to both non-Kōkihi ngā Rito groups. Interestingly, the average number of support hours per referral for tamariki supported by other services was significantly higher at refuges that deliver Kōkihi ngā Rito than at those without this programme.

We observed a similar pattern in case notes, with Kōkihi ngā Rito tamariki having significantly longer case notes across the board, but non-Kōkihi ngā Rito tamariki also having significantly longer case notes when supported at a refuge that also delivers the Kōkihi ngā Rito programme. Within just the "high-risk" cohort of tamariki, not only was this significant but the effect size was 0.19, well over the 0.14 usually taken as a large effect size. Not only is the difference significant, but the effect is also quite substantial and likely meaningful.

Based on the service delivery model for the Kōkihi ngā Rito programme, it was unsurprising to find that those referrals are significantly longer, provide significantly more hours of support per referral day, and result in significantly longer written case notes. What was more surprising was the apparent influence of the mere presence of a Kōkihi ngā Rito programme at a refuge on the support hours and case note length of other services supporting tamariki at that refuge. Given the expectation that children not referred into Kōkihi ngā Rito when the programme is available would represent lower risk or lower need, it is particularly unexpected to see multiple signatures of greater service delivery to these children.

As a rising tide lifts all ships, we hypothesise that the development of a Kōkihi ngā Rito programme within a refuge may help equip all Advocates – even those not directly delivering the programme – with practices that lead to more engagement with tamariki, potentially driving unanticipated changes to overall service delivery. Trained Kaiārahi tamariki may demonstrate standards for support hours per day and thorough documentation practices, indirectly influencing other contract streams or inspiring practice change in their peers. Within these two service delivery measures, we may be seeing a ripple effect where the Kōkihi ngā Rito programme may instil or encourage practices that have a far-reaching impact at the refuges, influencing not just individual interactions between caseworkers and tamariki but the overall service delivery standards within a refuge.

Limitations

With more than 40 independent refuges across the country, the individual context of each refuge obviously factors strongly into their service delivery. These analyses do not account for potential confounding factors such as individual staff characteristics (e.g., personality, experience), difference in management styles across refuges, or internal expectations for documentation, which could also influence service delivery interactions and tamariki wellbeing. Additionally, the local community of individual refuges may influence the type and distribution of referrals received, which could also confound our measure of referral duration or influence the amount of support required (and subsequently delivered).

We also note that service delivery was inferred from proxy variables which may not capture the nuances of overall service delivery. Calculations of average days per referral, support hours per day, and the number of characters in tamariki case notes provide one objective and easily measurable view of service delivery, but the absence of direct or qualitative feedback from tamariki limits that view.

Finally, Kruskal-Wallis tests can be sensitive to sampling, often becoming significant at large samples even when differences are just tiny deviations from the null hypothesis. To mitigate this effect and avoid misinterpreting results, we considered effect size alongside significance.

Conclusion

These analyses highlight the significant difference in service delivery practice within the Kōkihi ngā Rito programme and indicate that the much longer duration of these referrals is important to correct for when exploring other differences. Tamariki in Kōkihi ngā Rito have longer referrals, receive more hours of support per day, and have more detailed case notes written about their care. These observations reinforce that the Kōkihi ngā Rito programme delivers service differently from other contracts.

This exploration also illuminates an interesting and unexpected finding: even those tamariki not supported by the Kōkihi ngā Rito programme receive more hours of support per day and have more detailed case notes when they are being supported at a refuge that also delivers Kōkihi ngā Rito. Whether this represents a ripple effect of staff learning and practice improvement, or whether other unobserved factors may be at play, we recommend this area as a particular priority for further investigation. Staff interviews and conversations could help to identify a proposed mechanism of impact, and analyses of wāhine data across the same service delivery measures and refuge groups would confirm whether this effect is seen more broadly than just for tamariki. Detailed investigation of specific caseworker datasets could also help to uncover whether the effect is limited to certain staff or contract streams.

Casenote word frequency analyses

Statistical methods

The final analysis of the Kōkihi ngā Rito data delved into the text of case notes recorded by NCIWR staff across the three tamariki cohorts. An initial set of words was developed within the NCIWR Research & Policy team, and categorised as Advocacy, Feeling, Whānau/Other, and Violence (these words are provided in the Appendix). All case notes were scanned and the occurrence of each word in each case note was counted. From this data, several measures were calculated: the percentage of case notes containing each word, the average number of occurrences of each word within a case note, and the percentage of referrals where each word was used in at least one case note. These measures were selected to ensure that the observed differences in referral duration and number of case notes were controlled for.

Two-way ANOVA analyses were conducted to compare the effects of the two independent variables—tamariki cohort and word category—to identify whether they influenced each dependent measure and whether there was any interaction between these influences.

Interpretation of findings

The two-way ANOVA showed that there was a significant difference between tamariki cohorts represented in the independent variable KNRgroup in relation to average number of occurrences of advocacy, feeling, violence, and other words within case notes ($p=<0.001$). Additionally, there was a significant difference between those four-word categories ($p=<0.001$), and an interaction between the two variables KNRgroup and Category.

The average case note for a Kōkihi ngā Rito tamaiti contained almost twice as many Feeling words and Other words compared to the average case note for a non-Kōkihi ngā Rito child at any refuge. Additionally, non-Kōkihi ngā Rito tamariki supported within a refuge that also delivers Kōkihi ngā Rito were much more likely to have Advocacy words in their case notes—almost 3x more than those at a non-Kōkihi ngā Rito refuge, and a roughly 30% increase compared to the Kōkihi ngā Rito cohort. Violence terms were far less frequent across all three cohorts.

The bulk of words in the Other category relate to whānau members—mum, dad, brother, sister, friends, love, whānau, cousin—and the Kōkihi ngā Rito programme works specifically to support tamariki with relationships and feelings. Based on expectations of practice differences, it was unsurprising to see the much larger average occurrence of those words within Kōkihi ngā Rito tamariki.

The marked difference between cohorts observed in Advocacy words was particularly interesting. A deeper dive into the underlying data showed that the word “lawyer” was responsible for much of this difference—it occurs 25x more frequently in case notes for non-Kōkihi ngā Rito tamariki at a refuge delivering the programme, and 8x more frequently than in Kōkihi ngā Rito children. Close behind was the word “court”, which occurs almost 10x more frequently in the non-Kōkihi ngā Rito tamariki compared to those at a refuge that does not deliver Kōkihi ngā Rito, and 2x more frequently compared to the Kōkihi ngā Rito tamariki. This suggests a clear external confounder that is influencing the word choice within case notes and warrants conversation and investigation with refuge managers and staff. Is there a reason why children at a refuge delivering Kōkihi ngā Rito, but who are being supported by other services, should require so much more legal support? Whether this reflects local communities, refuge practice, the types of contract streams at certain refuges, or something else that happens to align quite coincidentally with the distribution of Kōkihi ngā Rito programmes, remains unclear. Further exploration could focus on individual refuges within these three tamariki cohorts to identify whether the signature is coming from one area in particular and consider wāhine case notes to investigate whether a similar pattern is observed. Conversations with subject matter experts would help to identify potential mechanisms behind this observed relationship in the data.

Limitations

These analyses are based upon words selected by subject matters experts, but this process could have introduced biases or unintentional omissions. Subsequent analyses extracting high-frequency words directly from the case notes themselves could help to mitigate this potential limitation.

Any analysis of free text is limited by the subjectivity that is inherent in its capture; Advocates or Kaiārahi tamariki may differ in how they document events, feelings, and observations. Variability in note-taking style, language used, and focus areas could result in inconsistent data and potential over- or under-reporting of key themes. Certain words or phrases (such as "whānau" or "violence") might be used more frequently in some cases depending on the worker's perception or focus area, which could introduce bias. For example, workers trained in trauma-informed care might emphasise emotional language more than those focused on procedural or legal aspects.

We also note that some case notes may be missing important information, especially in high-pressure situations where staff may not have time to document everything. Inconsistent documentation practices across different staff or refuges could limit the accuracy and completeness of the data, leading to skewed results. Addressing these limitations through triangulation with other data sources, such as interviews, setting clear documentation standards, and using mixed-methods approaches could help mitigate the impact of these issues on the analysis.

Table 10. Results of two-way ANOVA for KNRgroup and Category, for average number of occurrences

	Type III Sum of Squares	df	Mean Square	F	p-value	η^2_p
KNRgroup	**4,992.27**	**2**	**2,496.14**	**195.42**	**<.001***	**0.01**
Category	**14,569.18**	**3**	**4,856.39**	**380.19**	**<.001***	**0.03**
KNRgroup x Category	**2,844.22**	**6**	**474.04**	**37.11**	**<.001***	**0.01**
Error	437,925.33	34,284	12.77			

Bold indicates statistical significance at p=0.05

A guide to interpretation:

Type III sum of squares

This value represents the unique contribution of each factor to explaining the variance in the dependent variable, after accounting for the other factors and their interaction; a larger value indicates that this factor explains more of the variance. The value of 4,992.27 for the KNRgroup compared to 14,569.18 for Category indicates that the category of the word (Advocacy, Feeling, Violence, or Other) predicts about 3x as much of the variance in occurrence as does the tamariki cohort.

df

This is the degrees of freedom, which represents the number of independent variables that can vary within each factor. Calculated as the number of categories minus 1, df = 2.

Mean square

The mean square is the sum of squares (SS) divided by the degrees of freedom (df) for each factor or interaction term and represents the average amount of variation explained by each factor or interaction. A larger value indicates that more variation is explained by this factor.

F

The f-statistic compares the variance explained by each factor to the unexplained variance (or error). The larger the value, the more likely that the observed differences between groups are due to meaningful differences rather than random error.

p-value

The p-value helps to assess the significance of the f-statistic. A p-value of <0.001 is well below the 5% significance level, so we reject the null hypothesis; these data provide evidence that the tamariki cohort does significantly influence the average number of occurrences of these words within case notes.

η^2_p

This statistic, pronounced "partial eta squared," is a measure of effect size used with ANOVA tests. It quantifies the proportion of variance explained by each factor, after correcting for the other factors. Typically, a value of 0.01 is interpreted as a small effect, 0.06 is a medium effect, and 0.14, a large effect. Effect sizes help to provide practical interpretation of results alongside statistical significance.

Table 11. Bonferroni post-hoc tests for Category, for average number of occurrences

Category comparison	Mean difference	Std error	t	p-value
Advocacy <> Feeling	**0.17**	**0.055**	**3.18**	**0.009***
Advocacy <> Whānau	**-0.4**	**0.055**	**-7.34**	**<.001***
Advocacy <> Violence	**1.35**	**0.055**	**24.74**	**<.001***
Feeling <> Whānau	**-0.57**	**0.055**	**-10.52**	**<.001***
Feeling <> Violence	**1.18**	**0.055**	**21.56**	**<.001***
Whānau <> Violence	**1.75**	**0.055**	**32.08**	**<.001***

Bold indicates statistical significance at p=0.05

Table 12. Bonferroni post-hoc tests for KNRgroup, for average number of occurrences

KNRgroup comparison	Mean difference	Std error	t	p-value
KNR <> nonKNR at KNR refuge	**0.32**	**0.07**	**4.63**	**<.001***
KNR <> nonKNR refuge	**0.83**	**0.042**	**19.77**	**<.001***
nonKNR at KNR refuge <> nonKNR refuge	**0.5**	**0.073**	**6.84**	**<.001***

Bold indicates statistical significance at p=0.05

A guide to interpretation:

Mean difference

This represents the difference in average scores between two groups. A positive mean difference means the first group has a higher mean than the second, while a negative value means the second group has a higher mean. The value of 0.17 indicates that Advocacy word frequency has an average score that is 0.17 units higher than that of the Feeling words.

Std error

The standard error measures how much variation there is in the estimated mean difference due to sampling variability. A smaller value like 0.055 suggests the mean difference between Advocacy and Feelings frequencies is estimated more precisely.

t (test statistic)

The test-stastic is a standardised value that tells us how large the observed mean difference is relative to the standard error. A higher absolute value of 3.18 suggests a stronger difference between the occurrence frequencies of Advocacy and Feelings words.

p (p-value)

This tells us the probability that the observed difference occurred by chance, if there were actually no real difference between the groups. Since the Bonferroni method adjusts for multiple comparisons, the reported p-value is already corrected, and the value of 0.009 indicates the difference in frequency of Advocacy and Feelings words is unlikely to be due to chance.

Conclusion

Patterns of word use in tamariki case notes reflect service delivery practices within the Kōkihi ngā Rito programme compared to other services. Tamariki supported by Kōkihi ngā Rito have case notes with almost double the occurrences of feeling-related words and mentions of whānau members compared to other tamariki. These observations demonstrate the effective implementation of the Kōkihi ngā Rito programme. Designed to support tamariki with feelings and relationships, this distinctive focus and model of care are even reflected in the day-to-day, operational documentation by Kaiārahi tamariki.

Interestingly, the tamariki supported by other services at Kōkihi ngā Rito refuges have a dramatically higher frequency of case notes mentioning “lawyer” or “court”, suggesting the presence of a confounder external influence that warrants further discussion or investigation.

Risk assessment frequency analyses

Statistical methods

All risk assessments completed for wāhine between January 1, 2018 and July 31, 2024 were extracted to provide a descriptive summary overview of the composition and distribution patterns of risk items specifically affecting tamariki. This summary was overlaid with data regarding the presence and identities of tamariki within the household. No inferential statistical tests were performed.

Two different versions of the risk assessment were in use across this time period, so child-related questions were mapped according to logic in the Appendix into these six child-related risk categories:

- Children threatened with harm
- Children taken or kept from mum
- Children physically harmed
- Mum verbally abused in front of children
- Mum physically harmed in front of children
- Mum forced to do something sexual in front of children

Results

Within the time period, 20,921 risk assessments were completed for 15,856 unique wāhine, of whom 10,308 (65%) additionally provided identifiable tamariki details (at a minimum, name and date of birth) to be recorded in Recordbase. Those tamariki details amount to 23,112 individual children of mothers experiencing family violence, who are known to NCIWR. Of those tamariki, 4,479 (19%) have previously been supported in some way by NCIWR, and 156 (<1%) have participated in the Kōkihi ngā Rito programme.

Another 5,548 (35%) wāhine had a completed risk assessment, but did not provide tamariki details to be recorded. Of those, however, 2,804 (51%) answered Yes to at least one of the child-related risk questions, indicating that a child has been present at some point. This represents at least another 2,804 tamariki who may be experiencing family violence.

Together, this represents 25,916 known tamariki who may be experiencing family violence, just 4,479 (17%) having received support from NCIWR. On average, one wahine experiencing family violence represents an additional 1.63 tamariki experiencing the same.

Of the 15,856 wāhine who completed a risk assessment, 4,805 (30%) indicated that none of the child-related risks were present. These assessments connect back to 3,856 known tamariki whose mothers have sought support from NCIWR but who, according to these risk assessments, may not have yet been present or witness to the violence.

At the opposite end of the spectrum, 4,537 wāhine (29%) indicated the presence of all six child-related risks—and these assessments link to 7,471 individual known tamariki and likely another 591 from wāhine who did not provide tamariki details.

Of note, while it is least common for a wahine to report having been forced to do something sexual in front of children (4,426 wāhine have ever responded Yes to one of those questions), when that risk item is present, it is almost always present with all other five items. Nearly 85% of wāhine who reported having been forced to do something sexual in front of children also reported all five other child-related risk items.

Where only a single child-related risk was reported, it was most likely to be one of the two items related to the mother being physically or verbally abused in front of children—46% of single-risk responses were for verbal abuse of the mother, and 32% were for physical abuse of the mother.

Limitations

Risk assessment data is not collected for all wāhine accessing support at NCIWR, and is only required within the Community, Outreach and Residential services. The risk profiles of wāhine presenting to these contracts may not necessarily represent those who connect with NCIWR via other channels or contract streams.

We also know that the nature of the subject matter requires a great deal of trust, bravery, and vulnerability from women already experiencing harm. While Advocates are trained to collect this information with sensitivity, we acknowledge that the complexity and trauma of family violence forms the human lens through which these data points are provided. Other child-related harms may be present and described in narrative portions of the risk assessment, or wāhine may chose not to disclose certain harms due to shame or other personal choices. Situations within a whānau can change dramatically, particularly as wāhine navigate an exit from violence, and these nuances may not all be captured in new risk assessment forms.

Conclusion

These findings indicate a high prevalence of child-related risks reported by wāhine experiencing family violence, with at least one child-related risk being reported in 70% of all risk assessments conducted by NCIWR. Extrapolation based on the roughly 15,000 wāhine that NCIWR supports every year suggests that more than 25,000 tamariki could be reached through their mother's engagement with NCIWR—but that currently, fewer than 1 in 5 of those tamariki is likely to receive direct support.

Based on these data, we conclude that the potential reach for a programme like Kōkihi ngā Rito is large, and that the ready availability of identifiable tamariki details and risk profiles could facilitate access, engagement, and prioritisation.

Appendix

Technical methods

Data extraction and cleansing

All data was extracted from a single Azure SQL Server database for Recordbase, the client management system used by NCIWR, using t-sql code. Initial scripts were written by the senior data scientist to extract just Kōkihi ngā Rito data for further downstream analysis by other members of the team, and then subsequent transformation and consolidation scripts were written by multiple members of the data science team. All code was subject to internal peer review, and each data extract was conducted independently and cross-checked with colleagues to confirm identical database records were retrieved.

An initial scan of Kōkihi ngā Rito data indicated that several tamariki were missing key data elements, namely either a second My Star™ collection or a clear link to a mother client file (where risk assessment data was located). To facilitate a more complete dataset, a spreadsheet of data gaps was provided to MSD to send to the team at NCIWR to review. Where information was present on the tamariki client record in narrative format, they were able to update the record; for example, where a mother was mentioned by name in the case notes but a relationship had not been created between the two client files, the NCIWR data team added that missing data point. This process resulted in 4 additional tamariki records being cleansed for inclusion, increasing the Kōkihi ngā Rito cohort size used for My Star™ analyses from 67 to 71.

Whānau composition data

Relationship data connecting individual client files was used to establish mother-child relationships between individual client files. Whānau connections are complicated, and the dataset representing them is equally so. In Recordbase, individual relationships can be created between files, and those relationships operate in both directions. For example, in the screenshot following, the example client file Ashley Baker is a child of Mary and John, with siblings Thomas, Chloe and Emma.

Mary can be recorded as the mother of Ashley, or Ashley can be recorded as the child of Mary, or both can be recorded simultaneously. In practice, the choice usually depends on the order in which clients are entered into the software, so our code checked for all possible relationship types that could represent a parent-child connection (where the parent was female, to exclude father records) and then ran the same check in the opposite direction. Based on these multiple passes and the potential for a single relationship to be recorded multiple times in different ways, the final set of data required deduplication to produce the exhaustive list of (recorded) parent-child relationships.

Early work also included sibling relationships in these whānau reconstructions, but subsequent investigation revealed that these specific relationship types are not consistently used. To approximate the sibling count for whānau composition, we counted the number of unique tamariki that a child's mother was also related to; for example, Ashley's mother Mary is also recorded as the mother of Thomas, Chloe, and Emma, so the whānau composition for Ashley is 4.

Direct Relationships New

Linked Person	Relationship type	Primary Person	
Mary Smith	Relative / Parent/Matua to	Ashley Baker	Edit
John Smith Baker	Relative / Parent/Matua to	Ashley Baker	Edit
Thomas Baker	Relative / Sibling to	Ashley Baker	Edit
Chloe Baker	Relative / Sibling to	Ashley Baker	Edit
Emma Baker	Relative / Sibling to	Ashley Baker	Edit
Ashley Baker	With Mum of	Mary Smith	
Ashley Baker	Sibling of	Chloe Baker	
Ashley Baker	Sibling of	Emma Baker	
Ashley Baker	With Mum of	John Smith Baker	
Ashley Baker	Sibling of	Thomas Baker	

Address data

Address data was extracted for all tamariki cohorts to facilitate closer matching, based on the expectation that the deprivation index would be a better indicator of similarity across refuges. These individual addresses were processed using the Addressfinder API to return statistical area 2 (SA2), which was then used to identify the deprivation index. Not all addresses were able to be verified at first pass, and some required manual correction (for example, where an apartment number had been concatenated to the street number, or where an Auckland postcode was typed as 600 instead of 0600). This work was only performed for the cohorts involved in the PSM evaluations, although we continued to use deprivation index to provide insights into the larger cohorts; the much lower hit rate of deprivation index for those report pages is due simply to the lack of manual cleansing and post-processing, rather than any difference in actual data capture.

Outcomes and assessment data

Outcomes tool and risk assessment data were retrieved with additional t-sql code. Both items are natively collected in an xml schema, so the data processed to the reporting database is parsed but not normalised. Therefore, these datasets require significant additional work to unpivot and convert to appropriate data types. Additionally, the risk assessment data was present in two separate xml schema corresponding to the separate versions of the tool; these were extract separately, restricted to just the questions of interest (those relating to children), and then mapped together using the following logic.

Risk assessment mapping

Questions from the legacy risk assessment and the new risk assessment were mapped into single response categories according to the logic below. Where multiple questions correspond to a single final category, any affirmative response was taken as an indicator of the presence of that risk. For example, we code a Yes response to any one of these three questions on the legacy risk assessment as a Yes to *Children threatened with harm*:

- Has the perpetrator ever threatened harm to children?
- Has the perpetrator threatened to take your children?
- Has the perpetrator threatened to kill your children?

Legacy risk assessment question	New risk assessment question	Combined into
Has the perpetrator abused you in front of your children?	Have they harmed you in front of children?	Mum physically harmed in front of children
Has the perpetrator taken your children as a threat?	Have they taken your children, or threatened to?	Children taken or kept from mum
Has the perpetrator ever caused harm to children?	Have they physically harmed children?	Children physically harmed
Made you have sex in front of the children?	Have they forced you to do anything sexual when children could see or hear it?	Mum forced to do something sexual in front of children
Made you have sex with the children present or so they could hear?		
Has the perpetrator ever threatened harm to children?	Have they threatened to kill or hurt your children?	Children threatened with harm
Has the perpetrator threatened to take your children?		
Has the perpetrator threatened to kill your children?		
Tell your children negative things about you?	Have they put you down, belittled you, or verbally abused you in front of children?	Mum verbally abused in front of children

Statistical methods

Propensity score estimation

To estimate propensity scores, we conducted a logistic regression analysis where the treatment variable was regressed on the covariates, and the estimated propensity scores were then used to match treated and control participants. Univariate analyses were conducted with the covariates ethnicity, age, gender, sibling in KNR, risk, refuge, and days in service to determine relevance. Keen to avoid overfitting, we discarded risk, refuge, days in service, and sibling in KNR because their inclusion did not influence post-matching balance[54,55,56]. All variables are defined in the table on the following page.

Propensity score matching

We employed the `teffects psmatch` command in Stata to perform propensity score matching. This command implements nearest neighbour matching on an estimated propensity score, which is not to be confused with nearest-neighbour estimator for average treatment effect, implemented in the `teffects nnmatch` command.

Confusingly, the term "nearest neighbour" is used to refer to two different but related concepts in PSM and in machine learning or general statistical matching. In our context, "nearest neighbour matching" refers to a specific algorithm within PSM, which paired each treated unit (Kōkihi ngā Rito tamaiti) with one or more control units (a tamaiti supported by other services) based on the smallest difference between their propensity scores (i.e., the closest "neighbour" in terms of the propensity score). Propensity scores estimate the probability of receiving treatment, based on observed covariates, and the matching was based on those propensity scores – hence the general methodology name of Propensity Score Matching.

We used nearest neighbour matching **with replacement**, which means that a control unit can be matched with more than one treated unit. The alternative, without replacement, means that once a control unit is matched to a treated unit, it cannot be used again for other matches. Given the limited size of the tamariki dataset, we selected with replacement to minimise sample loss while maximising the statistical power of our tests.

Outside of PSM, nearest neighbour matching is a more general technique used in various areas of statistics, data science, and machine learning, and refers to a broader concept of identifying the closest data points based on distance metrics, commonly used in classification, regression, or clustering. For clarity, we matched based on propensity scores (hence, PSM), but the choice of which tamaiti to select as a match when comparing propensity scores was made using the nearest neighbour algorithm.

Results were validated by assessing balance between groups after matching, and a caliper size of 0.5 was selected based on visual diagnostics of the balance charts. The command was written as follows:

```
teffects psmatch (outcome_variable) (treatment_variable
covariate1 covariate2 ...)
```

where `outcome_variable` represents the dependent variable (generally the change in a specific wellbeing domain), `treatment_variable` indicates the treatment status (1 or 0 for KNR or nonKNR), and the covariates are those previously identified in the propensity score estimation: age, gender, and ethnicity.

Difference-in-Differences calculations

Initial DiD analyses were conducted in both Stata and R, with the calculations replicated in Microsoft Excel for ease of sensitivity analyses. This involved calculating averages for pre- and post-periods across KNR treatment and nonKNR control groups, computing pre-post differences, and taking the difference of those differences to estimate the treatment effect, calculating the multiple regression using the LINEST function. This same technique was replicated in PowerBI DAX calculations using the LINESTX function, and values crosschecked across all methods.

Statistical tests

All additional statistical tests – Kruskal-Wallis, Dunn-Bonferroni, and two-way ANOVA – were computed using DataTab software, with calculations rebuilt in PowerBI for visualisation. The Levene test of variance equality and programmatic histograms and QQ plots were reviewed within the software to justify test selection.

Propensity scoring variable definitions

Variable	Definition
KNR	A dummy variable that equals 1 for tamariki in KNR and 0 otherwise
Age	Age of tamaiti is created by subtracting the accepted date (into service) of the tamaiti from the date of birth, dividing by 365 and then using the integer of the number i.e 11.6 = 11 as a child isn't classified as 12 until they reach 12 years of age. This is a continuous variable and tamariki aged 5 to 12 inclusive are included.
Gender	A dummy variable is created for gender equal to 1 for female and 0 for male.
Ethnicity	The Statistics NZ definition for ethnicity is used. Due to the small sample size of Pacific tamariki in the KNR group only we combine Māori and Pacific into one group. Ethnicity is therefore a dummy variable that equals 1 for Māori and Pacific and 0 for all other (primarily European).
Date My Star™ was completed	A variable containing the day, month and year the My Star™ was completed. If the data was entered retrospectively, the date for which the star was actually completed is used.
Star order	All My Star™ measures completed for a child were put in order of the date when they were completed; this variable equals 1 on the date the first star was completed.
Average outcome	A continuous variable that is an average of the answers to each of the domains, 1 to 8 was created (if an answer = 0, it was changed to missing as a 0 answer indicates that domain wasn't completed).
Average star change	This variable is created by first selecting the first star (date) in the sequence (if tamariki have more than one star) and then the last star (date) in the sequence. The sequence is created by date where 1 is the first date completed. The first is subtracted from the last to create a continuous variable for the Average star change for each tamariki.
Previous Service Episodes	A dummy variable is created equal to 1 if a tamaiti had a previous service episode and 0 otherwise
Sibling	A dummy variable is created to indicate whether a tamaiti has a sibling in KNR or the control group service equal to 1 if yes or 0 if no
Risk Variable	If there is more than one risk variable for each tamaiti mum, the one closest to the service episode is used.
Days in service	This is a continuous variable that indicates how many days each tamaiti spent in service
Time between stars	This is a continuous variable that indicates how many days occurred between the first and last My Star™ reading.
Reading for each Domain	A continuous variable is created for each Domain, 1 to 8 for the first and last My Star™ readings.
Domain change	The difference between first and last My Star™ reading for each domain is calculated

Casenote word categories

The words and categories provided for case note analysis are as follows:

Violence terms:

- Kill
- Stab
- Assault
- Fight
- Hiding
- Hurt
- Punch
- Manipulate
- Coerce (coercion)
- Violence
- Abuse

Feelings terms:

- Guilt
- Afraid / fear / scared
- Worried / worry
- Anxiety / anxious

Advocacy/need depth terms:

- Lawyer
- Court
- Police
- Risk
- Safety / safety plan
- Money / financial
- Feelings
- Evidence
- School
- OT / FGC

Whānau/other terms:

- Dad
- Mum
- Partner
- Friends
- Love
- Whānau
- Cousin
- Friend

References

References

1. UN Development Fund for Women. (2011). Progress of the world's women: In pursuit of justice. Retrieved from: https://www.unwomen.org/en/digital-library/publications/2011/7/progress-of-the-world-s-women-in-pursuit-of-justice.
2. Superu. (2017). What works for children exposed to family violence? Retrieved from: https://thehub.sia.govt.nz/resources/what-works-for-children-exposed-to-family-violence/.
3. Martin, J., Langley, J., & Millichamp, J. (2006). Domestic violence as witnessed by New Zealand children. The New Zealand Medical Journal, 119(1228).
4. Fergusson, D.M., & Horwood, L.J. (1998). Exposure to interparental violence in childhood and psychosocial adjustment in young adulthood. Child Abuse & Neglect, 22(339-357).
5. New Zealand Police (2023). Annual report 2022-2023. Retrieved from: https://www.police.govt.nz/sites/default/files/publications/annual-report-2022-2023.pdf.
6. New Zealand Police (2022). Police Statistics on Homicide Victims in New Zealand 2007 – 2021: A Summary of Statistics about Victims of Murder, Manslaughter, and Infanticide. Retrieved from: https://www.police.govt.nz/sites/default/files/publications/homicide-victims-report-2022.pdf.
7. Mueller, I., & Tronick, E. (2019). Early life exposure to violence: developmental consequences on brain and behavior. Frontiers in Behavioral Neuroscience, 13(156).
8. Abel, K., Heuvelman, H., Rai, D., Timpson, N., Sarginson, J., Shallcross, R., Mitchell, H., Hope, H., & Emsley, R. (2019). Intelligence in offspring born to women exposed to intimate partner violence: a population-based cohort study. Wellcome Open Research, 4(107).
9. Koenen, K. C., Moffitt, T. E., Caspi, A., Taylor, A., & Purcell, S. (2003). Domestic violence is associated with environmental suppression of IQ in young children. Development and Psychopathology, 15(2), 297-311.
10. Artz, S., Jackson, M. A., Rossiter, K. R., Nijdam-Jones, A., Géczy, I., & Porteous, S. (2014). A comprehensive review of the literature on the impact of exposure to intimate partner violence on children and youth. International Journal of Child, Youth and Family Studies, 5(4), 493-587.
11. Kiesel, L. R., Piescher, K. N., & Edleson, J. L. (2016). The relationship between child maltreatment, intimate partner violence exposure, and academic performance. Journal of Public Child Welfare, 10(4), 434-456.
12. Cunningham, A., & Baker, L.L. (2007). Little eyes, little ears: How violence against a mother shapes children as they grow. Ontario, Canada: Centre for Children & Families in the Justice System, London Family Court Clinic, Inc. Retrieved from: https://publications.gc.ca/site/eng/313567/publication.html.
13. Howell, K. H., Barnes, S. E., Miller, L. E., & Graham-Bermann, S. A. (2016). Developmental variations in the impact of intimate partner violence exposure during childhood. Journal of Injury &Violence Research, 8(1), 43-57.
14. Harold, G. T., & Sellers, R. (2018). Annual research review: Interparental conflict and youth psychopathology: An evidence review and practice focused update. Journal of Child Psychology and Psychiatry, 59(4), 374-402.
15. Ingram, K. M., Espelage, D. L., Davis, J. P., & Merrin, G. J. (2020). Family violence, sibling, and peer aggression during adolescence: Associations with behavioral health outcomes. Frontiers in Psychiatry, 11(26).
16. McTavish, J. R., MacGregor, J. C., Wathen, C. N., & MacMillan, H. L. (2016). Children's exposure to intimate partner violence: An overview. International Review of Psychiatry, 28(5), 504-518.
17. Koss, K. J., & Gunnar, M. R. (2018). Annual Research Review: Early adversity, the hypothalamic–pituitary–adrenocortical axis, and child psychopathology. Journal of Child Psychology and Psychiatry, 59(4), 327-346.
18. Teicher, M. H., & Samson, J. A. (2016). Annual research review: Enduring neurobiological effects of childhood abuse and negelct. Journal of Child Psychology and Psychiatry, 57(3), 241 – 266.
19. Anderson, S. L., & Teicher, M. H. (2004). Delayed effects of early stress on hippocampal development. Neuropsychopharmacology, 29, 1988-1993.
20. Vu, N. L., Jouriles, E. N., McDonald, R., & Rosenfield, D. (2016). Children's exposure to intimate partner violence: A meta-analysis of longitudinal associations with child adjustment problems. Clinical Psychology Review, 46, 25-33.
21. Artz, S., Jackson, M. A., Rossiter, K. R., Nijdam-Jones, A., Géczy, I., & Porteous, S. (2014). A comprehensive review of the literature on the impact of exposure to intimate partner violence on children and youth. International Journal of Child, Youth and Family Studies, 5(4), 493-587.
22. Holmes, M. R. (2013). The sleeper effect of intimate partner violence exposure: Long-term consequences on young children's aggressive behavior. Journal of Child Psychology and Psychiatry, 54(9), 986-995.

23. McLaughlin, K.A., Weissman, D., & Bitrán, D. (2019). Childhood adversity and neural development: a systematic review. Annual Review of Developmental Psychology, 1, 277–312.

24. Whittle, S., Dennison, M., Vijayakumar, N., Simmons, J.G., Yucel, M., Lubman, D.I., Pantelis, C., & Allen, N.B. (2013). Childhood maltreatment and psychopathology affect brain development during adolescence. Journal of the American Academy of Child and Adolescent Psychiatry, 52(9), 940–952.

25. Martin, J., Langley, J., & Millichamp, J. (2006). Domestic violence as witnessed by New Zealand children. The New Zealand Medical Journal, 119(1228).

26. Women's Refuge (2024). Annual report 2023-2024. Retrieved from: https://womensrefuge.org.nz/wp-content/uploads/2024/10/Annual-Report-2024-ONLINE.pdf.

27. Triangle Consulting Limited (2025). Empowerment Star™: The Outcomes Star for women who have experienced domestic violence. Retrieved from: https://www.outcomesstar.org.uk/using-the-star/see-the-stars/empowerment-star/.

28. Thorburn, N., & Arathoon, C. (2023). Re-thinking 'risk': using risk assessment data in family violence and safety work. Retrieved from: https://womensrefuge.org.nz/wp-content/uploads/2023/05/Rethinking-Risk.pdf.

29. Arathoon, C., Thorburn, N., Burgess, M., & Jury, A. (2023). Evaluation of Kōkihi ngā Rito: What Kōkihi ngā Rito can teach us about building safety for tamariki who have experienced family violence. Retrieved from: https://womensrefuge.org.nz/wp-content/uploads/2023/11/Womens-Refuge-Evaluation-of-Kokihi-nga-Rito-digital.pdf.

30. Arathoon, C., Thorburn, N., & Jury, A. (2021). Kids in the Middle. Retrieved from https://womensrefuge.org.nz/wp-content/uploads/2021/08/KITM-WHOLE-FINAL.pdf.

31. Triangle Consulting Limited (2025). My Star™: The Outcomes Star for children. Retrieved from: https://www.outcomesstar.org.uk/using-the-star/see-the-stars/my-star/.

32. MacKeith, J. (2011). The development of the Outcomes Star: a participatory approach to assessment and outcome measurement. Housing, Care and Support, 14(3), 98-106.

33. Mudiyanselage, S. B., Wanni Arachchige Dona, S., Angeles, M. R., Majmudar, I., Marembo, M., Tan, E. J., Price, A., Watts, J. J., Gold, L., & Abimanyi-Ochom, J. (2024). The impact of maternal health on child's health outcomes during the first five years of child's life in countries with health systems similar to Australia: A systematic review. PloS one, 19(3), e0295295.

34. Ministry of Social Development (2022). Maternal health and children's socio-emotional and cognitive development: New evidence from the Growing Up in New Zealand study. Retrieved from: https://www.msd.govt.nz/documents/about-msd-and-our-work/publications-resources/research/children-and-families-research-fund/maternal-health-and-children-s-socio-emotional-and-cognitive-development.pdf.

35. Clapham, B., Breheny, M., Reweti, A., Severinsen, C., & Ware, F. (2024). Missed Opportunities for Addressing Maternal Mental Health: A Thematic Analysis of Mothers' Experiences of Using the Well Child Tamariki Ora Service in Aotearoa NZ. Health & Social Care in the Community, 2024(1), 5890641.

36. Sawilowsky, S. S., & Blair, R. C. (1992). A More Realistic Look at the Robustness and Type II Error Properties of the t-test to Departures from Population Normality. Psychological Bulletin, 111(2), 352-360.

37. Zimmerman, D. W. (1998). Inaccuracy of the Distribution of the Sample Mean and Misuse of the t-Test When Sampling from a Nonnormal Population. Journal of Experimental Education, 67(1), 55-65.

38. Lumley, T., Diehr, P., Emerson, S., & Chen, L. (2002). The Importance of the Normality Assumption in Large Public Health Data Sets. Annual Review of Public Health, 23(1), 151-169.

39. Schmider, E., Ziegler, M., Danay, E., Beyer, L., & Bühner, M. (2010). Is It Really Robust? Reinvestigating the Robustness of ANOVA Against Violations of the Normal Distribution Assumption. Methodology, 6(4), 147-151.

40. Cohen J. (1988). Statistical Power Analysis for the Behavioral Sciences. New York, NY: Routledge Academic

41. Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. Journal of Economic Literature, 47(1), 5-86.

42. Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. Biometrika, 70(1), 41-55.

43. Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. Journal of Economic Surveys, 22(1), 31-72.

44. Abadie, A. (2005). Semiparametric difference-in-differences estimators. The Review of Economic Studies, 72(1), 1-19.

45. Angrist, J. D., & Pischke, J.-S. (2008). Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press.

46. Mills, D. A., Thom, K., Maynard, A., Meehan, C., Kidd, J., Newcombe, D., & Widdowson, D. (2015). Meeting the housing needs of vulnerable populations in New Zealand. University of Auckland.

47. Gibb, K., & McGreal, S. (2018). Housing policies in New Zealand: A response to homelessness and family violence. Land Use Policy, 76, 712-719. doi:10.1016/j.landusepol.2018.03.031.

48. Klein, L. B., Chesworth, B. R., Howland-Myers, J. R., Rizo, C. F., & Macy, R. J. (2021). Housing Interventions for Intimate Partner Violence Survivors: A Systematic Review. Trauma, Violence, & Abuse, 22(2), 249-264. https://doi.org/10.1177/1524838019836284

49. Pawson, H., & Davidson, J. (2020). Housing and family violence: What can be done? Housing Studies, 35(1), 66-83. doi:10.1080/02673037.2020.1797667

50. Duncan, M., & Scott, L. (2022). The intersection of housing and family violence in New Zealand: A literature review. Building Better Homes, Towns and Cities.

51. Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. Review of Educational Research, 42(3), 237-288. https://doi.org/10.3102/00346543042003237

52. Lix, L. M., Keselman, J. C., & Keselman, H. J. (1996). Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance F test. Review of Educational Research, 66(4), 579-619.

53. Blanca, M. J., Alarcón, R., Arnau, J., Bono, R., & Bendayan, R. (2017). Non-normal data: Is ANOVA still a valid option? Psicothema, 29(4), 552-557. https://doi.org/10.7334/psicothema2016.383

54. Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. Biometrika, 70(1), 41-55.

55. Austin, P. C. (2011). "An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies." Multivariate Behavioral Research, 46(3), 399-424.

56. Imbens, G. W., & Rubin, D. B. (2015). "Causal Inference in Statistics, Social, and Biomedical Sciences." Cambridge University Press.

[Her Kaiārahi Tamariki] built her confidence back up, built her trust in the world. If your dad can do that to you, then everyone could do that... [This programme] has made her realise that he is not the rule—he is the exception to that rule. Most people are genuinely good people. It made her look at her own circle too—"actually, Mum is on my side, grandma and grandad are on my side". [Her Kaiārahi Tamariki] gave her that confidence to look at her own circle and realise there was one really bad egg in our group, but there are 100 people that are pushing and helping us and are there when we need them.

- Mum of Kōkihi ngā Rito child

Tūtohi

Wild Bamboo