Moving the Needle for At-Risk Children?

An Assessment of the Impact of the Stand for Children Service

Children's Services Tū Māia Whānau
A world strong with children

12th February 2017
tdb.co.nz

TDB Advisory Limited
L5, Wakefield House
90 The Terrace
P.O. Box 93
Wellington
New Zealand

Tel (+644) 934 8740
Email: info@tdb.co.nz

Acknowledgements:

We would like to kindly thank Dr Sarah Berry and Dr Lisa Underwood of the *Growing Up in New Zealand study* and Dr Ian Lambie of the University of Auckland for their helpful comments, as well as the staff at Stand Children's Services' Christchurch service centre for graciously hosting us and patiently answering our questions.

Principal contacts for this report:

Name: Phil Barry
Email: phil.barry@tdb.co.nz
Tel: (+6421) 478 426

Name: Tom Stannard
Email: tom.stannard@tdb.co.nz
Tel: (+6427) 800 8988

Disclaimer

Table of contents

1. Executive summary 1
2. Introduction 4
3. Stand Children's Services and the Stand for Children Service 5
4. Evidence and review of children at risk of poor future life outcomes 7
   - 4.1 Treasury's characteristics of "at-risk" youth 7
   - 4.2 Adverse Childhood Experiences study 9
   - 4.3 Growing Up in New Zealand cohort study 10
   - 4.4 Other large-scale longitudinal studies undertaken 10
   - 4.5 Summary of risk factors identified and studied 12
5. Behaviour and environment as "at-risk" indicators 14
   - 5.1 Strengths and Difficulties Questionnaire (SDQ) 14
   - 5.2 Environmental risk and behavioural profiles - is there a link in NZ? 18
   - 5.3 SDQ summary 19
6. The fiscal impact of at-risk children in the long-term 20
   - 6.1 NZ Treasury's estimates of the fiscal cost of at-risk children 20
   - 6.2 Estimates of the economic cost of maltreatment 22
   - 6.3 Summary of value and at-risk children 23
7. The Stand for Children Service's clients 24
   - 7.1 Demographic profiles 24
   - 7.2 Client behavioural profiles 32
   - 7.3 Summary of Stand's client profiles 37
8. The impact of the Stand for Children Service on behaviour profiles 38
   - 8.1 Average scores 38
   - 8.2 Change of distributions for Total Difficulties scores 42
   - 8.3 Improvement proportions and magnitudes 46
   - 8.4 Summary of the Service's impact 47
9. Fiscal saving and economic benefit estimates for the Stand for Children Service 48
   - 9.1 Producing a long-term positive financial outcome for the government 48
   - 9.2 Summary of fiscal indications and the effect of the Service 51
10. An alternative approach to estimating fiscal savings with uncertain outcomes 52
   - 10.1 Defining outcomes as distributional and uncertain 52
   - 10.2 The log Normal distribution and Geometric Brownian Motion 53
   - 10.3 Outcome estimation approach with uncertainty 54
   - 10.4 Estimating the distribution 55
   - 10.5 Estimating the distributional effect of a shift in trajectory 56
   - 10.6 Summary and possible implications for Stand and its funding for the Service 58
11. Recommendations for further work 59
   - 11.1 Data matching in the IDI of Stand's clients and their outcomes 59
   - 11.2 Longer-term assessment of the impact of the Service 59
   - 11.3 Other possible approaches to consider for assessment of interventions 60
   - 11.4 Summary of further work recommendations 61
12. Conclusions 62

Appendix 1: The intervention logic for Stand Children's Services 64

Appendix 2: Data-cleaning process 65

Appendix 3: SDQ assessments of Stand's clients by age and year 66

Appendix 4: Full breakdown of the normal to abnormal distributions of Stand's clients 67

Appendix 5: Supplementary analysis of average scores using a cohort approach 70
Appendix 6: Full breakdown of change directions and magnitudes 71
Appendix 7: More detailed outline of Monte Carlo estimation process 78

List of figures

Figure 1: Table 9 from NZ Treasury 2016 8
Figure 2: SDQ scoring and classification system 15
Figure 3: Table 4 from Stringaris & Goodman 2013 17
Figure 4: Table 14 from GUiNZ Vulnerability Report 2 18
Figure 5: Table 10 from NZ Treasury 2016, cost by specific risk factor 21
Figure 6: Table 12 from NZ Treasury 2016, cost by average number of risk factors 22
Figure 7: MSD findings of substantiated abuse by region 25
Figure 8: Heat map from NZ Deprivation Index (pp 33-34 NZDep2013) 26
Figure 9: The Service's clients by DHB identification 26
Figure 10: Average age of the Service's clients by year 28
Figure 11: Distribution of the Service's clients by age and year 28
Figure 12: The Service's gender ratio over time 29
Figure 13: Admission rates to Care & Protection and Youth Justice services by gender . 30
Figure 14: Ethnicity of the Service's clients by year 31
Figure 15: Distribution of the Service's clients' school decile since 2000 31
Figure 16: Stand Total Difficulties behaviour profile scores 36
Figure 17: Stand self-responses vs. GUiNZ behaviour profile scores 44
Figure 18: Stand self-responses vs. GUiNZ behaviour profile scores 45
Figure 19: Stand self-responses vs. GUiNZ behaviour profile scores 45
Figure 20: Stand's number of accepted referrals per year 48
Figure 21: Graphical representation of trajectory methodology in a point-estimate setting 52
Figure 22: Implied distribution of Treasury's numbers 55
Figure 23: Implied distribution for the one-risk-factor group 56
Figure 24: Implied distribution for one-risk-factor group with altered trajectory 57
Figure 25: Implied distributions compared 57
Figure 26: Stand's intervention logic 64
Figure 27: Observed frequency of Stand's SDQ responses by age 66
Figure 28: Observed frequency of Stand's SDQ responses by year 66

List of tables

Table 1: Catchment areas of Stand and service-centre location 24
Table 2: Age of the Service's clients 27
Table 3: Gender of the Service's clients 29
Table 4: Ethnicity of the Service's clients since 2000 30
Table 5: Average SDQ scores for full sample 32
Table 6: Classification of average SDQ scores 32
Table 7: Total observations used in each average SDQ score 33

Table 8: Distribution of full sample classifications for Total Difficulties SDQ 34
Table 9: GUiNZ-SDQ risk profile 35
Table 10: GUiNZ behaviour profile scores by environmental risk 35
Table 11: Stand Total Difficulties behaviour profile scores 36
Table 12: Average SDQ scores through the intervention process 38
Table 13: Absolute changes in SDQ scores 39
Table 14: SDQ behaviour classification through the intervention process 40
Table 15: Observations used in calculating average SDQ scores 41
Table 16: Distribution of SDQ behaviour profile through intervention 42
Table 17: Change in distribution of SDQ behaviour profile through intervention 43
Table 18: Stand behaviour profile scores throughout intervention 43
Table 19: Total Difficulties score observed directional shifts 46
Table 20: Total Difficulties score observed average percentage changes 47
Table 21: Recap of risk factors and NZ Treasury estimates of direct cost to NZ 49
Table 22: Implied cost saving by shifting a child the equivalent of one NZ Treasury risk factor 50
Table 23: Implied cost saving by shifting a child two NZ Treasury risk factors 50
Table 24: Detailed pre-intervention SDQ behaviour profile breakdown 67
Table 25: Detailed post-intervention SDQ behaviour profile breakdown 68
Table 26: Detailed six-month follow-up SDQ behaviour profile breakdown 69
Table 27: Average Total Difficulties scores using different cohort approaches 70
Table 28: Impact score direction shifts 71
Table 29: Impact score magnitude changes: averages and standard deviations 71
Table 30: Total Difficulties score direction shifts 72
Table 31: Total Difficulties score magnitude changes: averages and standard deviations 72
Table 32: Pro-Social score direction shifts 73
Table 33: Pro-Social score magnitude changes: averages and standard deviations 73
Table 34: Peer score direction shifts 74
Table 35: Peer score magnitude changes: averages and standard deviations 74
Table 36: Hyperactive score direction shifts 75
Table 37: Hyperactive score magnitude changes: averages and standard deviations 75
Table 38: Conduct score direction shifts 76
Table 39: Conduct score magnitude changes: averages and standard deviations 76
Table 40: Emotion score direction shifts 77
Table 41: Emotion score magnitude changes: averages and standard deviations 77

List of equations

Equation 1: Behaviour profile score 35
Equation 2: Geometric Brownian Motion 53
Equation 3: GBM solution 54
Equation 4: SDQ Value Added Score 60

1. Executive summary

Stand Children's Services (Stand) is a not-for-profit organisation providing services for at-risk children. The Stand for Children Service (hence forth referred to as 'the Service') is an intervention aimed at children who are at risk of poor future life outcomes. The Service is segmented into two components:

1. a wrap-around service; and
2. therapeutic care and education.

The first component involves the Service working with the family and community to adapt the environment of the child to maximise the likelihood of a long-term positive outcome. The second component takes children who have been referred to the Service and targets past trauma and current behavioural difficulties.

The purpose of this report is to assess the impact that the Stand for Children Service has on its clients, in terms of long-term life outcomes for the children and what that may mean in terms of financial savings for the New Zealand government over time. To undertake this study, we seek to address the following three questions:

1. is the Service targeting the right children?
2. is the Service having a positive impact on the children? and
3. what is that impact worth in terms of fiscal savings to the government over time?

In regards to the first question, the evidence indicates that Stand is overall successful in targeting children who are at higher risk of poor future life outcomes than usual. We find that the children referred to the Service:

- appear to be at higher risk, measured by substantiated finding of abuse, or have higher general deprivation;
- are on average approximately 50:50 NZ European and Maori, with an increasing percentage of Maori and other ethnicities in recent years;
- are almost 2:1 males: females, consistent with Child Youth and Family (CYF) and Youth Justice (YJ) admission rates;[1] and
- are more likely to come from schools with low decile ratings.

Behaviourally, we find that children referred to the Service:

- on average, exhibit borderline to abnormal total behavioural difficulties;
- appear to have behavioural distributions that are riskier than the average child; and

[1] http://www.cyf.govt.nz/about-us/key-statistics/kids-in-care.html#DistinctchildrenandyoungpeopleinthecustodyoftheChiefExecutive2

- on average, exhibit a behavioural profile comparable to the 'high environmental risk' cohort in the Growing Up in New Zealand (GUiNZ) study.[2,3]

In answer to the second question, our analysis indicates that on average the Service has a positive impact on the behavioural profiles of the children referred to it. More specifically we find that:

- while not all children benefit, there is evidence of meaningful positive behavioural change, on average, across the behavioural difficulties metrics used by Stand;[4]

- prior to the intervention (pre-intervention), parents and teachers rate between 50%-60% of the children using the service as having a behavioural profile consistent with the child having abnormal behavioural difficulties. After the intervention (post-intervention) and at a six-month follow-up assessment, this falls to approximately 40%;

- on average, of those who had an improved behavioural profile following Stand's Service, the improvement was approximately one third from the base line; and

- in the parent and teacher assessments of behaviour, on average, the children referred to the Service have a change in behaviour profile that shifts from being above the profile expected of the high environmental risk group, to below it (effectively moving the average child from a profile consistent with a high-risk environment to a profile consistent with a medium-risk environment). This may indicate a positive material shift in likely future life outcomes.

The available information does not permit us to definitively answer the third question (the fiscal impact of Stand) at this stage. Using estimated future costs of children in at risk environments produced by the New Zealand Treasury,[5] we estimate that if the Service has a positive[6] impact on approximately one in six of the children referred to it, the government will "break-even" (ie, generate future fiscal savings that recover the current cost of funding). If it can exceed this standard, it will generate a positive financial return for the government.

Using an alternative approach, we ran a simulation to compare the distribution of future costs to the government of at-risk children with the distribution of expected future costs of these children following the positive behavioural profile changes following the children's contact with the Service. We assume that the improved behavioural profile changes (refer Section 7) correlate with a 33% decrease in the likelihood of a poor-outcome trajectory for these children. The average cost to the government (up to age 35) of a child who is exposed to one Treasury risk factor is $100,300 with an inferred standard deviation of $26,826. Our simulation yields an expected average future cost of $72,089 with a decreased standard deviation of $19,273 indicating an overall tighter distribution of likely outcomes and associated costs.[7]

---

[2] As discussed in Section 5, the Service uses the internationally recognised Strengths and Difficulties Questionnaire to assess behavioural profile of the children it serves.
[3] Risk profiles are analysed in "Vulnerability Report 2: Transitions in exposure to vulnerability in the first 1000 days of life", Growing Up in New Zealand Report 6, 2015.
[4] Defined by the Growing Up in New Zealand study, http://www.growingup.co.nz/
[5] Refer Sections 4 and 6 below.
[6] By "positive" we mean decreasing the child's risk profile by effectively one NZ Treasury risk factor on average, as discussed in Sections 6 and 9 of this report.
[7] Standard deviations were calculated from the resulting simulated outcome samples.

Finally, this report presents other metrics that could be applied in assessing and comparing possible interventions and highlights, in particular, the value that could be added by integrating the Stand database with the micro data available through the Integrated Data Infrastructure (IDI) at Statistics New Zealand. By defining and analysing a comparison group it will make it much easier to assess the true benefits of the Service.

2. Introduction

TDB Advisory Ltd (TDB) has been engaged by Stand Children's Services (Stand) to carry out an independent assessment of its program, the Stand for Children Service (the Service) and the outcomes that the Service delivers. Section 3 of this report provides a brief overview of Stand and the Service. Section 4 surveys some previous work that has been conducted on identifying risk factors that can lead to poor future life outcomes for children. Section 5 discusses behavioural risk characteristics and presents one link between behaviour and risk of future poor life outcomes. Section 6 outlines recent work on the long-term costs to the government associated with risk factors of poor future outcomes in children. Section 7 presents the demographic and behavioural profiles of the children that Stand serves. Section 8 then provides an analysis of the impact that Stand has on its clients through its intervention process. Section 9 derives our estimate of the impact Stand would need to have to "break even" over time from a fiscal (ie, government financial) perspective. Section 10 presents a new approach to estimate the fiscal savings of Stand's services, in an attempt to represent uncertainty of future outcomes. Section 11 presents recommendations for future work and finally, Section 12 provides some concluding comments.

3. Stand Children's Services and the Stand for Children Service

Stand Children's Services (Stand) is a not-for-profit organisation providing services for at-risk children and their families. Stand has a long history of operating in New Zealand, tracing its origins back to the Children's Health Camp movement that began in 1919. Over the years, Stand's services, governance and operating models have evolved substantially.[8]

Stand is a charitable trust governed by a board of trustees. Stand's services are largely government funded and the Board provides the assets that make the delivery of services possible. In the 2015/16 year, Stand's total revenue was $22.6M, with 95% of its revenue coming from government sources and the rest largely from interest and income from investments and fundraising. As at 30 June 2016, Stand held assets of $55M and employed around 450 staff providing nationwide services. Stand's national office is in Wellington and service delivery is structured around seven regions (Northern, Auckland, Midland, East Coast, Central, Christchurch and Southern).

In this report, we focus on the Stand for Children Service (the Service) provided by Stand. The Service is an intervention aimed at children who are at risk of poor future life outcomes, and is segmented into two components:

1. a wrap-around service, and
2. therapeutic care and education.

The Service is referred approximately 1,600 children per year and receives funding from the Ministry of Social Development and the Ministry of Education amounting to approximately sixteen million dollars per year (indicating funding of approximately ten thousand dollars per child treated).

The focus of the Service, under agreement with the government, is to provide a national response to protect the most vulnerable children, aged five to 12 years, from further maltreatment or exposure to chronic trauma, and to support their recovery from that exposure and enhance their wellbeing.

The Service is therefore focused on and structured to achieve:

- a reduction in the risk to and an increase in the protection of Children, through an intensive wrap-around multi-systemic approach, to make changes within the family/whānau, school and community, through the provision of an Intensive Family Service; and
- supporting the Child's recovery and improving their wellbeing, through the provision of therapeutic care and education.

The Service is "trauma informed", meaning that service delivery is grounded in an understanding of, and responsiveness to the impact of trauma that emphasises physical, psychological and emotional safety for children and creates opportunities for them to

[8] For a brief history, refer http://standforchildren.org.nz/our-whakapapa.

rebuild a sense of control and empowerment and rediscover a love of relating and learning. This approach delivers the essential building blocks of recovery and wellbeing, ensuring children are given the best opportunity to recover from trauma leading to a better life at home, school and in their community and, growing up to be capable adults. The following short and long term outcomes for children and their families/whānau are extracted from Stand's service model intervention logic (illustrated in Appendix 1).

By providing the Service, Stand intends to meet both short and long-term outcomes.

Short-term outcomes (service results) include:

- children remaining in home and at school;
- improved family/whānau functionality;
- improved child mental, emotional and physical health;
- improved child and family/whānau social problem solving;
- reduced antisocial and at-risk behaviour from the child; and
- improved school readiness and cognitive abilities.

Long-term outcomes (population-based) include:

- decrease of children in statutory care or increase in Home for Life placements;
- improved child health and educational achievement;
- decrease in family/whānau violence, abuse and neglect;
- reduction in youth crime involvement; and
- reduced alcohol and substance abuse.

4. Evidence and review of children at risk of poor future life outcomes

This section presents a review of some of the many studies that have identified several common risk factors in children's lives that can lead to poor life outcomes. Section 4.1 presents the findings of specific studies undertaken by the NZ Treasury[9,10] on risk factors of at-risk children. Section 4.2 presents the Adverse Childhood Experiences study (ACEs).[11] Section 4.3 presents the Growing Up in New Zealand (GUiNZ) study a recent New Zealand based longitudinal study.[12] Section 4.4 provides a brief overview of other relevant domestic and international studies and Section 4.5 summarises.

4.1 Treasury's characteristics of "at-risk" youth

There has been increasing attention in recent years from the New Zealand Government on matters concerning social investment and assessing the costs associated with long-term social outcomes. A series of working papers have been produced by several government agencies that attempt to utilise Statistics New Zealand's Integrated Data Infrastructure (IDI) which contains micro level data obtained from several governmental agencies. This approach matches data on observations and outcomes of individuals, such as contact with Youth Justice (YJ), Child Youth and Family (CYF), and Corrections, within a cohort and tracks each individual through time, consistent with a longitudinal cohort framework.

A December 2015 working paper[9] prepared by the Analytics and Insights team within the NZ Treasury, in conjunction with Ministry of Education (MoE), is a key study utilising the integrated administrative data (a subset of the IDI). The paper identifies several characteristics among young people that are predictive of future outcomes. The characteristics identified in the Treasury study are:

- contact with government agencies (Child, Youth and Family);
- demographic characteristics;
- location;
- characteristics of caregiver; and
- early outcomes relating to education, corrections, welfare and health.

As subjects progress through adolescence into early adulthood (15 years - 24 years), the analysis of the administrative data ties the poor future outcomes more directly to individual contact with the benefit, corrections and health systems.

[9] "Using Integrated Administrative Data to Identify Youth Who Are at Risk of Poor Outcomes as Adults", Keith McLeod, Robert Templeton, Christopher Ball, Sarah Tumen, Sarah Crichton and Sylvia Dixon, New Zealand Treasury, December 2015.
[10] "Characteristics of Children at Greater Risk of Poor Outcomes as Adults", New Zealand Treasury Analytics and Insights Team, February 2016.
[11] https://www.cdc.gov/violenceprevention/acestudy/
[12] http://www.growingup.co.nz/

Building on the administrative data analysis for present risk factors predicting poor future outcomes, a February 2016 report[8], also prepared by the Analytics and Insights team within the NZ Treasury, matches some of the risk characteristics identified to the costs incurred by the government on individuals. The risk factors identified are:

- substantiated finding of abuse or neglect;
- mostly supported by welfare benefits since birth;
- parent with a sentence history; and
- mother with no formal qualifications.

Of the 510,351 children sampled, 8.3% had a substantiated finding of abuse or neglect, 12.8% were mostly supported by welfare benefits since birth, 17% had a parent with a sentence history and 11.3% had a mother with no formal qualification. Figure 1 presents Table 9 from the Treasury's paper, with the key summary statistics for each of the four risk indicators.

Figure 1: Table 9 from NZ Treasury 2016

Table 9: Characteristics of children aged 6 to 14 with each indicator present

	Total	Indicator			
		Substantiated finding of abuse or neglect	Mostly supported by welfare benefits since birth	Parent with a sentence history	Mother with no formal qualifications
Number of children	510,351	42,543	65,184	86,949	57,417
Percentage of children	100.0	8.3	12.8	17.0	11.3
	Percentage				
Gender					
Male	51.2	51.1	50.7	51.4	51.3
Female	48.8	49.0	49.3	48.6	48.7
Ethnicity					
Asian	9.4	2.9	3.1	1.9	3.5
European	52.6	27.7	23.2	25.5	40.2
Māori	26.0	54.7	59.1	59.4	42.4
Other	2.0	1.2	1.8	0.8	1.1
Pacific	10.0	13.6	12.9	12.3	12.8
Indicator					
Substantiated finding of abuse or neglect	8.3	100.0	29.3	30.1	19.7
Mainly supported by benefits since birth	12.8	45.0	100.0	40.2	28.1
Parent has a community or custodial sentence history	17.0	61.6	53.7	100.0	37.5
Mother has no formal qualifications	11.3	26.5	24.7	24.8	100.0
Selected characteristics					
Parental characteristics					
Received income support at the time of the birth	19.8	59.6	75.0	54.6	42.0
Had a previous or current gang affiliation	1.6	8.4	6.9	8.9	4.1
Mother was single at birth	20.4	52.9	64.9	46.6	36.7
Mother was teenager at birth	5.9	17.3	19.9	17.9	13.1
Safety					
Had a police family violence referral to CYF	9.6	58.9	33.9	33.0	20.4
Notified to CYF	20.0	99.8	58.2	56.8	40.2
Had a injury-related hospitalisation	15.2	24.4	21.0	20.8	18.5
Mother smoked around the time of the childs birth	9.8	24.5	25.3	25.2	20.7
Health					
Had an ambulatory sensitive hospitalisation	22.2	33.9	34.4	33.2	31.3
Low weight at birth	6.1	9.3	8.3	8.1	7.9
Belonging					
Changed address at least once a year on average	0.9	5.2	3.1	3.7	2.7

The paper establishes a link between these characteristics and the projected future outcomes for these children, as well as the estimated costs to the New Zealand Government (discussed in Section 6 of this report). The poor future outcomes identified in this paper are:

- increased likelihood of further contact with CYF and YJ;

- lower school achievement;
- increased probability of receiving Sole Parent support and the benefit;
- higher likelihood of a community or custodial sentence; and
- increased probability of having a gang affiliation.

4.2 Adverse Childhood Experiences study

The Adverse Childhood Experiences study (ACEs) presents different risk characteristics for at-risk children. The ACE study has identified eight risk factors for a child who is more likely to experience poor life outcomes. The risk factors are:

- physical abuse;
- sexual abuse;
- emotional abuse;
- physical neglect;
- emotional neglect;
- mother treated violently;
- household substance abuse;
- household mental illness;
- parental separation or divorce; and
- incarcerated household member.

The study has linked these risk factors with a number of poor life outcomes, including but not limited to:

- future illicit drug abuse;[13]
- mood and anxiety disorders;[14]
- obesity;[15]
- mortality and many forms of morbidity, including:
    - autoimmune diseases;
    - liver diseases;
    - coronary diseases; and
    - pulmonary diseases.[15]

[13] "Childhood Abuse, Neglect, and Household Dysfunction and the Risk of Illicit Drug Use: The Adverse Childhood Experiences Study", Dube et al, Pediatrics, 2003, 111(3): 564-572.
[14] "Impacts of adverse childhood experiences on health, mental health, and substance use in early adulthood: A cohort study of an urban, minority sample in the U.S.", Mersky et al, Child Abuse Neglect, 2013, 37(11): 917–925.
[15] "Adverse Childhood Experiences and the Lifelong Consequences of Trauma" American Academy of Paediatrics, 2014.

4.3 Growing Up in New Zealand cohort study

The *Growing Up in New Zealand* study (GUiNZ) is an ongoing, New Zealand-based, longitudinal cohort study. It tracks the lives of approximately 7,000 children in New Zealand who were due to be born between 25 April 2009 and 25 March 2010 whose mothers were residents of Counties-Manukau, Waikato, or Auckland District Health Boards. The study has produced several reports that document its findings. A 2014 report[16] identifies several characteristics among children vulnerable to poor life outcomes at birth and throughout the first two years of life.

The specific risk factors identified by the 2014 GUiNZ report are:

- maternal depression;
- maternal physical wellbeing;
- maternal smoking;
- maternal age;
- maternal relationship status;
- maternal education;
- financial distress;
- deprivation area;
- unemployment;
- tenure living in social housing;
- income-tested benefit; and
- overcrowding.

The study develops the risk factors associated with children who are at increased vulnerability of poor developmental trajectories from previous longitudinal studies.

GUiNZ discourages isolating any one risk factor the study has identified when considering possible interventions. This is because the risk factors are linked only with very early outcomes and vulnerabilities (age 2 years) of children, and, risk factors appear to be clustered. Meaning that targeting one risk factor may not have any effect at all, if the other risk factors associated with it are not also addressed. The director of GUiNZ has said:

> *"Single risk factors (such as absolute or relative poverty) are commonly used to define early vulnerability, however targeting single risk factors for intervention(s) has limited capacity to minimise downstream adverse outcomes associated with that risk factor at a population level".*[17] *– Dr Susan Morton*

4.4 Other large-scale longitudinal studies undertaken

The Dunedin Multidisciplinary Health and Development Study is one of the older studies in New Zealand. Its cohort was born in 1972/1973. Many papers have been published

[16] "Vulnerability Report 1: Exploring the Definition of Vulnerability for Children in their First 1000 Days", Growing Up in New Zealand Report 4, 2014.
[17] "Vulnerability Report 1: Exploring the Definition of Vulnerability for Children in their First 1000 Days", Growing Up in New Zealand Report 4, 2014. p (iv).

over time on its findings. One recent study[18] which follows the cohort until age 32 demonstrates that (while noting caveats of lack of control for confounding factors) receiving the benefit has a statistically significant association with:

- lower family occupational status;
- having a mother who was young when she first became pregnant;
- low parental education;
- time in a sole parent family;
- multiple caregiver or residential changes;
- low family cohesion and high family conflict;
- physical abuse and sexual abuse.

Also, receiving the benefit was found to be associated with individual behavioural characteristics, including:

- socialised aggression;
- inattention;
- hyperactivity;
- conduct disorder;
- anxiety;
- antisocial behaviour;
- lower IQ;
- mental health problems; and
- lower self-esteem.

The Christchurch Health and Development Study is a longitudinal study of a birth cohort of 1,265 children born in urban Christchurch in mid-1977. This study has published more than 100 papers. It finds similar risk factors to the studies cited above, including:

- adverse family factors;
- sexual orientation and mental health; and
- depression and anxiety leading to unemployment and educational underachievement.[19]

Another finding from the study is that those exposed to more than 6 months of unemployment had rates of disorder that were 1.5 to 5.4 times higher than those who did not experience this time frame of unemployment.

The Avon Longitudinal Study of Parents and Children study (ALSPAC)[20] is from the UK, involves a cohort of approximately 10,000 children recruited in 1991 and 1992 and focuses heavily on health outcomes, such as indicators of autism. The ALSPAC study

[18] "Lifecourse factors associated with time spent receiving benefit in young adulthood: A note on early findings", David Welch, Moira Wilson, 2010.
https://www.msd.govt.nz/documents/about-msd-and-our-work/publications-resources/research/sole-parenting/lifecourse-factors-associated-with-benefit-receipt-summary-report.pdf
[19] "The Christchurch Health and Development Study: review of findings on child and adolescent mental health", David Fergusson and John Horwood, Christchurch School of Medicine, 2001.
[20] Cohort Profile: The Avon Longitudinal Study of Parents and Children: ALSPAC Mothers' Cohort, Fraser et al, International Journal of Epidemiology, 2013, 42(1): 97–110.

identifies similar risk characteristics to the GUiNZ study, such as age and education of mother, income, smoking during pregnancy and partner status.

Growing Up in Australia: The Longitudinal Study of Australian Children (LSAC) is another longitudinal cohort study. Like most of the other studies it tracks many children (over 10,000) in different waves recruited in 2003-2004 at different ages. As with many of the other cohort studies it focuses on the health and behaviour of children from birth. It also presents a list of other longitudinal studies undertaken to date.[21] It outlines 42 different longitudinal studies that, from a high-level overview, appear to identify risk factors very similar to those identified in the Treasury and GUiNZ reports.

4.5 Summary of risk factors identified and studied

This section of the report has presented a brief overview of a sample of relevant studies conducted into environmental risk factors for children that are correlated with poor future life outcomes for a child at different stages of development.

At a high-level, some common factors were:

- parental education levels and age;
- exposure to domestic violence;
- neglect and emotional abuse;
- socio-economic status and family financial situation; and
- close relationship to state services and corrections.

Some of the common life outcomes associated with children in at-risk environments have included an increased probability of:

- drug and alcohol abuse;
- being unemployed for longer amounts of time;
- heart, liver and other diseases;
- increased contact with corrections and other State services; and
- serious depression and other mental illnesses.

These findings have come from studies in a number of countries and concern children of different ages and stages in the developmental process. Each study may not be directly comparable to the service that Stand delivers and the clients referred to it, however there is a clear and robust finding that children in at-risk environments are more likely to have poor life outcomes.

It is also evident that many of the risk-factors commonly identified by these studies cannot realistically be altered by Stand. The service Stand provides cannot undo a case of abuse for a child, nor can it (meaningfully) give a child's mother a formal qualification. Furthermore, the risk factor identified in a study may be merely an indicator of a deeper and more systemic problem. A mother with no formal qualification may indicate that the mother herself grew up in an environment that didn't have the resources to afford the time and direct cost of education or an environment that didn't place any value on the

[21] http://www.growingupinaustralia.gov.au/resources/links.html

education process. In turn it may be more likely that the mother will pass this upbringing on to the child and the child will have less options and be more likely to have poor life outcomes. The most useful analysis will therefore not focus on whether Stand has been able to reduce or remove these risk-factors, but rather if Stand has been able to offset the damage caused by those risk factors or the systemic causes related to them on future outcomes for children.

5. Behaviour and environment as "at-risk" indicators

This section of the report presents the Strengths and Difficulties Questionnaire (SDQ), a widely-used behavioural assessment tool of children. The SDQ is an important metric used later in this report to determine the overall effectiveness of Stand. Section 5.1 presents a background of the assessment and some evidence of its effectiveness. Section 5.2 details recent use of the SDQ in *the Growing Up in New Zealand study* and Section 5.3 summarises.

5.1 Strengths and Difficulties Questionnaire (SDQ)

The Strengths and Difficulties Questionnaire (SDQ) has been used in many countries as an outcome-based assessment of behavioural symptoms.[22] The questionnaire is filled out by parents, teachers and children (who are old enough, from the age of 11). The SDQ has many metrics, all of which have been developed in an on-going nature through peer-reviewed academic studies since the SDQ's inception in 1999. As per its official website, the SDQ has almost 4,000 academic articles published on it from 102 countries. In New Zealand it has been utilised in the NZ Ministry of Health's (MOH's) 'B4 School Check' programme, where the SDQ is used to help identify where a child has/is going to have difficulties learning.[23]

The SDQ is broken into five sub-categories of symptoms - both positive and negative - that identify and track behavioural symptoms of an at-risk profile. These sub-categories are independent sections within the questionnaire and are aimed at quantifying:

- emotional symptoms;
- conduct problems;
- hyperactivity/inattention;
- peer-relationship problems; and
- pro-social behaviour.

The last factor, the pro-social behavioural score, represents the positive attributes of the child in his/her resilience to adverse experiences.

Each of the above five symptoms is assessed and scored out of 10. The four negative symptoms and problems scores (ie, the emotion, conduct, hyperactivity and peer-relation scores) are added together to obtain a Total Difficulties score (ranked out of 40). Each of the individual symptom scores and the Total Difficulties score are then classified into three categories: normal; borderline; and abnormal. This classification system has more recently been updated to a new four category approach: close to normal; slightly raised; high; and very high. The newer four category approach appears to not be as widely adopted and researched in the literature and for that reason, this report adopts the three-classification system.

[22] http://www.sdqinfo.org/
[23] http://www.health.govt.nz/our-work/life-stages/child-health/b4-school-check/b4-school-check-information-early-learning-services

Another metric developed to generate outcome-based measures of predictive behavioural aptitude and later life outcomes are the SDQ Impact scores. The Impact scores (scored out of ten) are based around rating the 'impact' on the home and school environment and social inclinations of the child as opposed to the more general behavioural characteristics of the traditional SDQ.

The SDQ scoring system (of normal, borderline or abnormal) for each of the constructs has determined, for an average random sample of children, to rate 80% of the children as normal, 10% borderline and 10% abnormal.[24]

The general SDQ scoring system is presented in Figure 2 below.

Figure 2: SDQ scoring and classification system

Table 3: Categorising SDQ scores for 4-17 year olds

	Original three-band categorisation			Newer four-band categorisation			
	Normal	Borderline	Abnormal	Close to average	Slightly raised (/slightly lowered)	High (/Low)	Very high (very low)
Parent completed SDQ							
Total difficulties score	0-13	14-16	17-40	0-13	14-16	17-19	20-40
Emotional problems score	0-3	4	5-10	0-3	4	5-6	7-10
Conduct problems score	0-2	3	4-10	0-2	3	4-5	6-10
Hyperactivity score	0-5	6	7-10	0-5	6-7	8	9-10
Peer problems score	0-2	3	4-10	0-2	3	4	5-10
Prosocial score	6-10	5	0-4	8-10	7	6	0-5
Impact score	0	1	2-10	0	1	2	3-10
Teacher completed SDQ							
Total difficulties score	0-11	12-15	16-40	0-11	12-15	16-18	19-40
Emotional problems score	0-4	5	6-10	0-3	4	5	6-10
Conduct problems score	0-2	3	4-10	0-2	3	4	5-10
Hyperactivity score	0-5	6	7-10	0-5	6-7	8	9-10
Peer problems score	0-3	4	5-10	0-2	3-4	5	6-10
Prosocial score	6-10	5	0-4	6-10	5	4	0-3
Impact score	0	1	2-10	0	1	2	3-10
Self-completed SDQ							
Total difficulties score	0-15	16-19	20-40	0-14	15-17	18-19	20-40
Emotional problems score	0-5	6	7-10	0-4	5	6	7-10
Conduct problems score	0-3	4	5-10	0-3	4	5	6-10
Hyperactivity score	0-5	6	7-10	0-5	6	7	8-10
Peer problems score	0-3	4-5	6-10	0-2	3	4	5-10
Prosocial score	6-10	5	0-4	7-10	6	5	0-4
Impact score	0	1	2-10	0	1	2	3-10

Figure 2 above shows the scoring system for the SDQ by construct.[25] Scoring and classification is not strictly consistent for the Total Difficulties score and its constructs. For example, a Total Difficulties score of 0-13 will classify a child as normal on that scale if the parent is the respondent, while a score of between 0-11 would classify a child as normal on the scale if the child is the respondent. This appears to be because, as stated above, the scores have been chosen to fit 80% into the normal, 10% into the borderline and 10% into the abnormal category, as defined on a population-based UK survey. It should be noted that self-reviews are only administered by children who are old enough to have the cognitive ability required (currently set at eleven years old).

[24] http://www.ehcap.co.uk/content/sites/ehcap/uploads/NewsDocuments/236/SDQEnglishUK4-17scoring-1.PDF
[25] http://www.ehcap.co.uk/content/sites/ehcap/uploads/NewsDocuments/236/SDQEnglishUK4-17scoring-1.PDF

5.1.1 Criticisms of the SDQ

A review of the use and application of the SDQ was carried out by the Centre for Person Centred Research at Auckland University of Technology (AUT) in 2014.[26] The AUT study found challenges with implementing the SDQ on a large scale, in particular it expressed concerns in the way the SDQ was communicated and administered. It demonstrates that the SDQ may be volatile in the accuracy of its representation across different ethnic and socio-economic groups. At the high-level, it outlines that in the Ministry of Health's 'B4 School Check' there appears to be some resistance from families to categorising their children in some settings (particularly from the Maori and Pacific Island community who were found to be adverse to putting their children in high-level classified groups). The AUT study recommends that if the SDQ is to be continued to be used, then there needs to be more communication of what it is and what it is measuring. Despite this, the SDQ is still used in the 'B4 School Check'. Our understanding of AUT's critique is that the SDQ is a good measure if it is understood properly.

Another recent study[27] on the validity of using the SDQ has found no bias in results across gender, ethnic group and socio-economic background, again indicating that the SDQ can be an objective and informative metric in the New Zealand setting when properly administered. This study utilised SDQ data administered to the GUiNZ cohort.

Another slightly older study presents a similar critique of the use of the SDQ, it again presents concerns with applying assessments such as the SDQ in an aggregate setting.[28] It expresses concerns over the rate of false positives of learning disabilities in children aged 4-5 years old, for which the SDQ has not been shown to be particularly well suited. Given that Stand's service is not aimed at diagnosing learning disabilities and is not primarily aimed at 4-5 year olds like in the 'B4 School Check', this may be less of a problem for the Service that Stand provides. This finding has also been indirectly challenged by D'Souza et al. (2017) which, as noted above, demonstrates the validity of SDQ for the GUiNZ cohort (aged 2 years old).

A further limitation to the SDQ is that the scoring system has been calibrated to roughly fit 80% of respondents as normal, 10% as borderline and 10% as abnormal for a UK population, and as such may not be fully transferrable to a New Zealand population. However, D'Souza et al. (2017) summarises SDQ data administered to the GUiNZ cohort and finds that of the entire population of data collected 78.9% of the children had Total Difficulties scores that were classified as normal, 11.3% had scores that classified them as borderline and 9.8% had Total Difficulties scores that classified them as abnormal.[29] This finding, although conducted on a sample of two year olds is a sample from New Zealand children.

---

[26] "A validation and norming study of the strengths and difficulties questionnaire in the New Zealand context", Centre for Person Centred Research, 2014. https://www.health.govt.nz/system/files/documents/publications/validation-norming-study-strengths-difficulties-questionnaire-nz-executive-summary-aug15.pdf

[27] Psychometric Properties and Normative Data for the Preschool Strengths and Difficulties Questionnaire in Two-Year-Old Children, S D'Souza et al, J Abnorm Child Psychol, (2017), 45:2, 345-357.

[28] "B4 School Report: A Critique of a Child Health Screening and Intervention Programme", Robert Miller, 2013, http://robertmiller-octspan.co.nz/octspan/wp-content/uploads/2013/12/B4School-Report-final.pdf

[29] Table 4, pg 351 of, Psychometric Properties and Normative Data for the Preschool Strengths and Difficulties Questionnaire in Two-Year-Old Children, S D'Souza et al, J Abnorm Child Psychol, (2017), 45:2, 345-357.

Overall, while the SDQ, like any measure in the social sciences, has its limitations, we consider it to be a robust measure that is internationally recognised and that is suitable for assessing the impact of Stand's Service.

5.1.2 *Strengths and Difficulties Questionnaire: prediction of future behaviour and correlation to child maltreatment*

A recent 2013 UK based study[30] has shown there is a positive and statistically significant relationship between SDQ scores and future contact with state services, self-harm, truancy and police contact from the baseline to the three-year follow-up. The study takes a sample of 4,479 children (52% male with an average age of 10.75 years old). Figure 3 below presents Table 4 from the published paper.

Figure 3: Table 4 from Stringaris & Goodman 2013

Table 4 Predictive validity: association between impact and psychosocial adjustment

		Reporting source		
	Predictors	Parent OR 95 % CI	Teacher OR 95 % CI	Youth OR 95 % CI
Outcome: service contact	Impact only	**1.56**	**1.69**	1.10 ns
		1.40, 1.74	1.49, 1.91	0.87, 1.41
	Impact adjusted for baseline total SDQ score	**1.19**	**1.36**	0.92 ns
		1.01, 1.41	1.09, 1.70	0.70, 1.23
Outcome: self harm	Impact only	**1.39**	**1.46**	**1.24**
		1.26, 1.55	1.30, 1.64	1.08, 1.43
	Impact adjusted for baseline total SDQ score	**1.24**	**1.37**	1.08ns
		1.07, 1.44	1.12, 1.67	0.92, 1.28
Outcome: truancy	Impact only	**1.42**	**1.50**	1.20 ns
		1.19, 1.71	1.20, 1.87	0.94, 1.53
	Impact adjusted for baseline total SDQ score	0.93 ns	0.88ns	0.95 ns
		0.70, 1.24	0.61, 1.27	0.70, 1.29
Outcome: police contact	Impact only	**1.28**	**1.44**	**1.46**
		1.15, 1.44	1.27, 1.63	1.28, 1.65
	Impact adjusted for baseline total SDQ score	1.01ns	1.16ns	1.12 ns
		0.86, 1.19	0.94, 1.42	0.96, 1.31

OR odds ratio; *CI* 95 % confidence interval. Odds ratios with confidence intervals are presented from logistic regression models with each of the outcomes as dependent variables and impact as an independent variable either unadjusted or adjusted for baseline total SDQ score

All findings in bold are significant ($p<0.05$)

Figure 3 presents the predictive power of the SDQ Impact scores and the Impact scores. The numbers in the table are Odds Ratios calculated from logit regressions where the SDQ scores are the dependent variable, and dummy variables for whether each child has come into contact with state services, has self-harmed, has been truant, or has been in contact with police in the following 36 months of the original SDQ being conducted, appear to make up the independent variables. It shows that the parent and teacher SDQ assessments are more informative than the self ("Youth" in the table) assessment, as indicated by the statistical significance (the bold numbers in the table are statistically significant at the 95% level). For context, the first reported Odds Ratio of 1.56 (the

[30] "The Value of Measuring Impact Alongside Symptoms in Children and Adolescents: A Longitudinal Assessment in a Community Sample", Argyris Stringaris & Robert Goodman, Journal of Abnormal Child Psychol (2013) 41:1109–1120.

parent response for the service contact outcome) means that higher SDQ scores are tied to a 1.56 times increase in the probability of service contact for a child in the sample in the 36 months following the initial assessment.

These findings indicate a clear relationship between SDQ scores and the future probability of a poor life trajectory. However, we are in some ways limited in how conclusive we can be about future life outcomes. Odds Ratios give an idea of probabilistic outcomes but if we want to think about how a change in SDQ (which is how we are wanting to assess Stand) changes the probability of outcomes for children, then we need the "marginal effects" from the regression. This changes the fitted coefficient of the logit regression into an intercept that can be interpreted in a similar way to an Ordinary Least Squares coefficient. To our knowledge this cannot be done without re-estimating the regression which is out of scope for the purposes of this project.

Another study has demonstrated that there is a positive and statistically significant correlation between SDQ scores and child maltreatment (and a constructed child maltreatment index. It has demonstrated a statistically significant "convergent validity", which is a positive correlation between the maltreatment index and the SDQ Total Difficulties scores, emotion scores, conduct scores, hyperactivity scores, peer problem scores and impact scores. That is, as the SDQ scores mentioned get larger (indicating a worse behavioural profile), the child in question has a higher maltreatment likelihood. Also, it shows a statistically significant "divergent validity" to the pro-social SDQ score, meaning the higher the pro-social score the more likely the child in question is to move away from the maltreatment index.

5.2 Environmental risk and behavioural profiles - is there a link in NZ?

The environment that a child is born into, over-time, is likely to affect his/her general behavioural profile. *The Growing Up in New Zealand study*, examines this relationship. Figure 4 below presents Table 14 from the Growing Up in New Zealand Vulnerability report 2. It shows the breakdown of SDQ profile by vulnerability risk group as defined by GUiNZ (and that we are interpreting throughout this report as a proxy for environmental risk).

Figure 4: Table 14 from GUiNZ Vulnerability Report 2

Table 14: Distribution of SDQ outcomes at two years amongst the antenatal vulnerability risk groups

	Vulnerability risk group - antenatal		
Total SDQ category	Low n (%)	Medium n (%)	High n (%)
Low risk	2270 (79.4)	1706 (62.8)	273 (36.9)
Borderline	351 (12.2)	486 (17.9)	141 (19.1)
Abnormal	238 (8.3)	523 (19.2)	325 (43.9)

In the study, the vulnerability of each child was classified based on data collected antenatally [31] by the defined environmental risk that the child would be born into. After two years of life, GUiNZ carried out an SDQ Total Difficulties assessment of the children

[31] Meaning before birth, during pregnancy.

(by the parent). The study reported that of those born into a defined low risk environment, 79.4% were classified as normal (specified as "low risk" [32] in the GUiNZ table), 12.2% were classified as having a borderline SDQ score and finally, 8.3% were classified as having an abnormal SDQ score.

In the medium environmental (or vulnerability) risk group, 62.8% were classified as normal by the SDQ, 17.9% were classified as having a borderline SDQ score and finally, 19.2% were classified as having an abnormal SDQ score.

Lastly, of those born into a high vulnerability risk environment only 36.9% were classified as normal, 19.1% were classified as having a borderline score and finally 43.9% were classified as having an abnormal SDQ score.

The GUiNZ study indicates a clear relationship between the environment a child is born into and his/her general behavioural profile at age 2. This is a finding that will help form and drive the analysis in Section 7 and Section 8 of this report (after the fiscal costs to the government of at-risk children has been reviewed).

5.3 SDQ summary

The evidence discussed in this section indicates a positive and statistically significant relationship between SDQ scores (a behavioural assessment tool) and future contact with state services, self-harm, truancy and police contact. There is also a link between the environment a child is born into and the child's behavioural profile and SDQ scores, as supported by *the Growing Up in New Zealand study*.

[32] GUiNZ has altered the terminology, perhaps because a classification of 'normal' is not the best classification approach, but we wish to stay consistent with the SDQ defined constructs.

6. The fiscal impact of at-risk children in the long-term

This section presents some of the work that has been conducted on the cost to the government associated with at-risk children. We focus heavily on two studies that have estimated the costs. Firstly (and most importantly), we consider the NZ Treasury's estimates of the average cost to the government of children given specific factors of a risky environment in early life (in Section 6.1). Secondly, we note estimates of the average economic cost of child maltreatment from a US-based study (in Section 6.2). Finally, Section 6.3 presents a summary of the findings.

6.1 NZ Treasury's estimates of the fiscal cost of at-risk children

Building on the administrative data analysis of risk factors predicting poor future outcomes, a February 2016 report prepared by the NZ Treasury[33] applies some of the risk characteristics identified to the costs incurred by the government on individuals. The risk factors identified (as noted in Section 4 above) are:

- substantiated finding of abuse or neglect;
- mostly supported by welfare benefits since birth;
- parent with a sentence history; and
- mother with no formal qualifications.

By matching observed outcomes from these risk factors at the micro level using the Statistics NZ IDI, the Treasury estimates the average cost expected to be incurred by the government for a child currently 0-5 years old and for a child currently 6-14 years old. Due to Stand's target clients being generally of school age, we focus on the Treasury's estimates for children between the ages of 6-14.

Figure 5 presents the table from the Treasury report of estimated costs for each risk characteristic identified by the Treasury. It shows that from the total pool of 6-14-year-old children sampled, the average cost to the government is $29,400 by the age of 21 and $67,700 by the age of 35. If a child has a substantiated finding of abuse or neglect this cost rises to $111,300 by the age of 21 and $222,300 by the age of 35. If the child has been mostly supported by welfare benefits since birth, the estimated cost becomes $103,300 by the age of 21 and $198,000 by the age of 35. If the child has a parent with a sentence history, the cost becomes $80,300 by the age of 21 and $167,300 by the age of 35. Finally, if the child has a mother with no formal qualification, the cost estimated by the Treasury using the IDI matching is $58,500 by the age of 21 and $122,400 by the age of 35. These cost estimates are in 2014 New Zealand dollars and are not discounted (ie, not in present value terms). If the costs were discounted they would be lower. On the other hand, however, they do not take into account the full fiscal costs as no allowance is made for lost tax revenue and the estimates only include the direct costs

[33] "Characteristics of Children at Greater Risk of Poor Outcomes as Adults", New Zealand Treasury Analytics and Insights Team, February 2016.

of certain interactions of the individual up to age 35 and are unlikely to include the full costs of the individual's interactions with the state.

Figure 5: Table 10 from NZ Treasury 2016, cost by specific risk factor

Table 10: Projected future outcomes for children aged 6 to 14 with each indicator present

	Total	Indicator			
		Substantiated finding of abuse or neglect	Mostly supported by welfare benefits since birth	Parent with a sentence history	Mother with no formal qualifications
Number of children	510,351	42,543	65,184	86,949	57,417
Percentage of children	100.0	8.3	12.8	17.0	11.3
Projected outcomes before age 21 (%)					
Further contact with CYF	18.3	75.6	51.7	48.4	34.8
Contact with Youth Justice	4.6	19.7	15.0	14.1	9.5
Did not achieve any school qualifications	17.7	42.7	38.6	35.4	27.6
Did not achieve a level 2 qualification	27.3	56.0	52.1	49.0	40.3
Received Sole Parent Support	5.2	15.8	15.2	13.4	9.7
Received a benefit for more than 2 years	7.4	26.9	22.6	19.6	14.4
Used mental health services or pharms	18.5	33.0	25.7	26.0	22.4
Received a community or custodial sentence	6.1	22.0	17.1	17.1	11.5
Received a custodial sentence	1.7	9.1	6.0	6.0	3.7
Has a current or previous gang affiliation	0.6	3.7	2.7	2.6	1.5
Projected outcomes when aged 25 to 34 (%)					
Received a community or custodial sentence	8.4	24.7	20.7	20.0	14.2
Received a custodial sentence	3.2	11.6	9.0	8.7	5.9
Spent more than 5 years receiving benefit	8.6	25.1	21.4	19.1	14.4
Projected average costs before age 21					
Average benefit costs when a child	18,500	52,300	72,100	47,400	35,900
Average care and protection costs	3,300	31,200	9,200	12,800	8,500
Average Youth Justice costs	600	3,500	2,000	2,100	1,300
Average total cost of CYF, YJ, benefits as a child	22,400	86,900	83,300	62,300	45,700
Average benefit costs	5,800	17,600	15,500	13,600	10,200
Average corrections costs	1,200	6,800	4,500	4,400	2,600
Average total cost*	29,400	111,300	103,300	80,300	58,500
Projected average costs before age 35					
Average adult benefit costs	37,700	100,200	88,900	80,300	61,300
Average corrections costs	7,700	35,100	25,700	24,800	15,300
Average total costs*	67,700	222,300	198,000	167,300	122,400

* includes CYF, YJ, child benefit costs

The costs included in the Treasury study that are likely to be incurred by the government on children with the four identified risk factors include benefit costs, care and protection costs, Youth Justice costs, and corrections costs, both when the child is growing up and in later life.

Unfortunately, it is not possible for us to directly tie the specific risk factors presented in the Treasury study above to Stand's clients without Stand's data being linked to the micro-data in the IDI.

Figure 6 presents Table 12 from Treasury's February 2016 report presenting the Treasury's estimates of the fiscal cost of the number of prevalent risk factors on average.

Figure 6: Table 12 from NZ Treasury 2016, cost by average number of risk factors

Table 12: Projected future outcomes for children aged 6 to 14 years by number of indicators present

	Total	Number of indicators					Two or more	Three or more
		None	One	Two	Three	Four		
Number of children	510,351	354,864	87,234	44,142	19,857	4,254	68,250	24,111
Percentage of children	100.0	69.5	17.1	8.6	3.9	0.8	13.4	4.7
Projected outcomes before age 21								
Further contact with CYF	18.3	7.9	27.7	51.9	73.3	83.7	60.1	75.1
Contact with Youth Justice	4.6	1.7	6.5	14.0	22.5	26.7	17.2	23.2
Did not achieve any school qualifications	17.7	11.9	23.1	37.2	47.0	52.7	41.0	48.0
Did not achieve a level 2 qualification	27.3	19.9	36.0	51.1	60.1	66.0	54.6	61.1
Received Sole Parent Support	5.2	2.5	7.9	13.7	19.1	21.0	15.7	19.4
Received a benefit for more than 2 years	7.4	3.3	10.6	20.7	30.3	35.3	24.4	31.1
Used mental health services or pharms	18.5	16.1	20.8	26.4	31.8	34.3	28.5	32.3
Received a community or custodial sentence	6.1	2.8	8.6	16.9	24.4	29.2	19.9	25.3
Received a custodial sentence	1.7	0.5	2.3	5.8	9.8	12.6	7.4	10.3
Has a current or previous gang affiliation	0.6	0.1	0.8	2.2	4.5	5.9	3.1	4.7
Projected outcomes when aged 25 to 34								
Received a community or custodial sentence	8.4	4.8	11.9	20.1	27.2	30.3	22.8	27.7
Received a custodial sentence	3.2	1.6	4.7	8.8	12.7	13.8	10.2	12.9
Spent more than 5 years receiving benefit	8.6	4.9	12.2	20.2	27.0	29.0	22.7	27.4
Projected average costs before age 21								
Average benefit costs when a child	18500	6900	32900	55700	67700	75700	60400	69100
Average care and protection costs	3300	400	3600	14100	24900	28100	18100	25500
Average Youth Justice costs	600	100	700	1900	3600	4900	2600	3800
Average total cost of CYF, YJ, benefits as a child	22,400	7,400	37,300	71,800	96,200	108,700	81,200	98,400
Average benefit costs	5,800	3,200	8,300	14,500	19,400	21,700	16,400	19,800
Average corrections costs	1,200	300	1,500	4,200	7,700	9,000	5,500	8,000
Average total cost*	29,400	10,800	47,100	90,500	123,300	139,300	103,100	126,100
Projected average costs before age 35								
Average adult benefit costs	37,700	23,200	52,900	84,300	108,700	119,500	93,600	110,600
Average corrections costs	7,700	2,700	10,100	24,200	40,000	47,200	30,300	41,300
Average total costs*	67,700	33,300	100,300	180,300	244,900	275,300	205,000	250,300

* includes CYF, YJ, child benefit costs

Figure 6 above shows that by the age of 35 the estimated average future cost to the government of a child with no risk factors is $33,300. If the number of risk factors increases to one risk factor, the expected fiscal cost increases to $100,300; two risk factors increase this cost to $180,300; three risk factors has an estimated cost of $244,900; and four risk factors has an estimated cost of $275,300. We use this data in Section 9 below to infer estimates of the fiscal consequences of the Stand for Children Service.

6.2 Estimates of the economic cost of maltreatment

As has been discussed in Sections 4 and 5 above, there is evidence from the literature that maltreatment of a child can be reflected in that child's behavioural profile. There has also been analysis of the economic costs of child maltreatment. One recent study[34] from the USA published in the Journal of Child Abuse & Neglect estimates that the present value of future costs associated with child maltreatment is approximately $210,012 (USD 2010) for non-fatal cases and $1.27m (USD 2010) for fatal cases. In the non-fatal case. The costs include $32,648 in childhood health-care costs, $10,530 in adult medical costs, $7,728 in child-welfare costs, $6,747 in criminal-justice costs, $7,999 in special-education costs and $144,360 in productivity losses. Productivity costs account for approximately 69% of the total estimated economic costs of maltreatment. Our report does not explicitly consider productivity losses as we are addressing (consistent with

[34] "The economic burden of child maltreatment in the United States and implications for prevention", Xiangming Fang, Derek S. Brown, Curtis S. Florence, James A. Mercy, Child Abuse & Neglect 36 (2012) 156-165.

Treasury and the methodology of the social investment framework) fiscal costs avoided from intervention. The fiscal costs estimated by the US study (US$65,652, or NZ$93,789 when converted at $0.7USD/NZD) are comparable in magnitude with the NZ Treasury's estimates of the average fiscal costs of a child who has one risk factor ($100,300).

6.3 Summary of value and at-risk children

This section has presented evidence that is available of the estimated fiscal costs associated with children who grow up in an at-risk environment. It has shown that:

- based on estimates from the NZ Treasury, a child with at-risk characteristics can cost the government between $100,300 and $275,300 on average (up to the age of 35), depending on the number of risk characteristics the child has;
- the costs likely to be incurred by the government include benefit costs, care and protection costs, Youth Justice costs, and corrections costs, both when the child is growing up and in later life; and
- there are important qualifications to these Treasury cost estimates. On the one hand, the estimates are not discounted and therefore overstate the costs in present value terms. On the other hand, they may understate the true fiscal costs as they do not include some fiscal items which would likely be correlated with at-risk children, such as foregone tax revenue. Further, the fiscal estimates are likely to understate the total economic cost associated with the risks factors. The US study suggests the fiscal costs are only around 30% of the full economic costs.

7. The Stand for Children Service's clients

This section addresses our first question in regards to the Stand for Children Service:

Is Stand targeting the right children?

In this section, we analyse the demographic and behavioural profiles of the children referred to the Service. Section 7.1 of this report breaks down the demographic profiles of the clients, analysing the regions the children come from, as well as their ages, gender, ethnicity and, school decile. Section 7.2 then presents the behavioural profiles of clients before intervention which we compare to the profiles of the children in the Growing Up in New Zealand (GUiNZ) cohort. Finally, Section 7.3 summarises our findings.

7.1 Demographic profiles

This subsection presents the demographic profiles of the Service's clients based on data provided by Stand. Firstly, we analyse the regions and catchment areas Stand focuses on. Secondly, we examine the ages of the referred clients, their gender, ethnicity and school deciles.

7.1.1 Client regions and possible risk by region

Table 1 presents the regions that Stand operates in and receives referrals from.

Table 1: Catchment areas of Stand and service-centre location

Name of Stand Region	Regional Service Centre	TLA Regions Covered
Northland	Whangarei	Northland
Auckland	Pakuranga	Auckland
Midland	Rotorua	Waikato, Bay of Plenty
East Coast	Gisborne	Gisborne (East Coast), Hawkes Bay
Central	Otaki	Taranaki, Manawatu-Wanganui , Kapiti-Horowhenua, Wellington, Marlborough, Tasman, Nelson
Christchurch	Christchurch	Canterbury, West Coast (and can include Marlborough and Nelson)
Southern	Roxburgh	Otago, Southland

Table 1 shows the region covered by the Service, the location of the Service's operation in that region (the location of the Service centre) and the source of the referrals (by Territorial Local Authority (TLA)) for that region.

To better gauge the general risk profiles by region, we consider data from the Ministry of Social Development (MSD), which provides the number of children and young people

with a substantiated finding of abuse by region[35] and the Deprivation Index, an indicator of deprivation, by region, developed by the Department of Public Health, Otago University.[36]

The MSD data on substantiated findings of abuse by region is depicted for the years 2014-2016 in Figure 7.

Figure 7: MSD findings of substantiated abuse by region

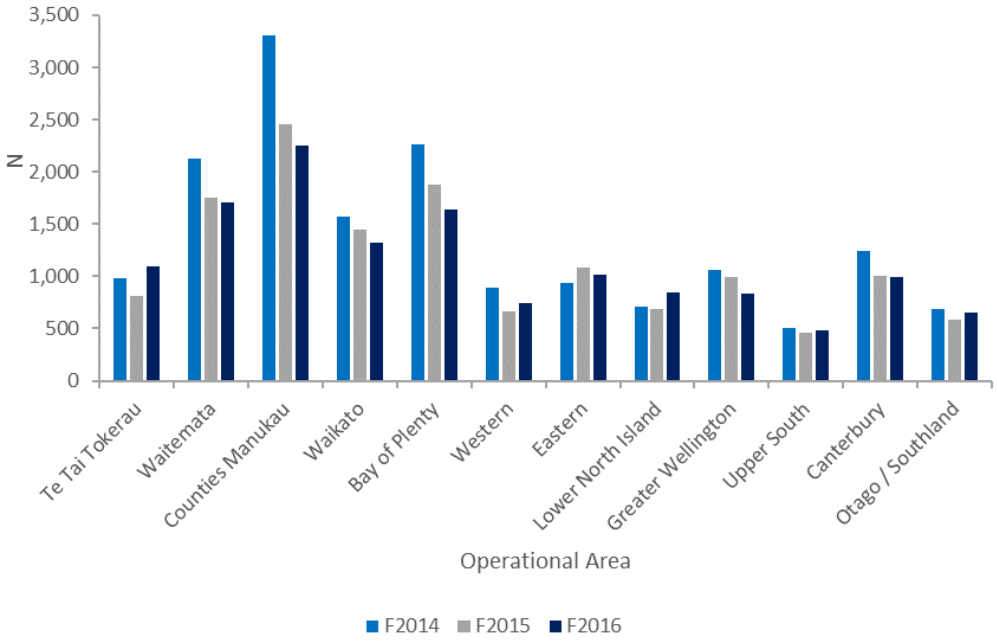


Figure 7 shows that for the South Island, the region with the highest number of children or young people with a substantiated finding of abuse is the Canterbury region, followed by the Otago/Southland region, with the Upper South Island having the lowest frequency. In the North Island, Counties Manukau has the highest frequency of substantiated abuse. This is followed by Bay of Plenty and Waitemata, then Waikato and Te Tai Tokerau Region (Northland). The lower and middle North Island appear to have low frequencies of substantiated abuse relative to the upper North and in most cases, lower than that of Canterbury.

[35] http://www.cyf.govt.nz/documents/about-us/key-statistics/findings-national-and-local-level-data-jun-2016.xlsx

[36] http://www.otago.ac.nz/wellington/otago069936.pdf

Figure 8 below present a heat map of the 2013 NZ Deprivation Index.

Figure 8: Heat map from NZ Deprivation Index (pp 33-34 NZDep2013)

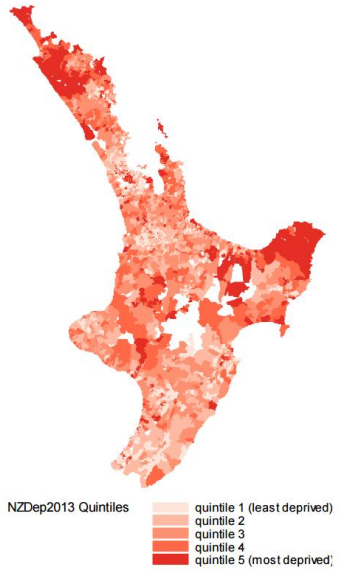


Figure 4: NZDep2013 distribution in the North Island of New Zealand

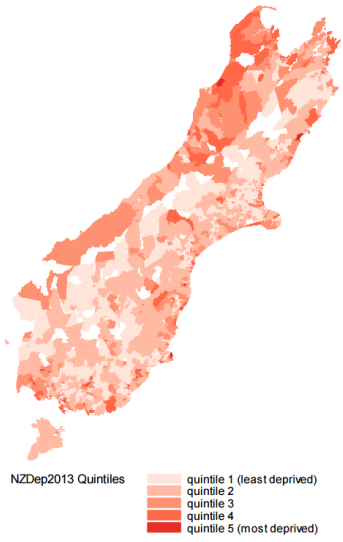


Figure 5: NZDep2013 distribution in the South Island of New Zealand

In Figure 8, darker red represents higher relative levels of deprivation. The map indicates that for the North Island, Gisborne/Bay of Plenty and Northland have the highest levels of deprivation. It also shows that the South Island has a considerably lower deprivation level than the North, and particularly the far North, which is consistent with the findings of substantiated abuse from the MSD data.

From the data, presented in Figure 7 and 8, it seems that risk profiles are likely to be higher in the Gisborne, Counties Manukau, Bay of Plenty and Waitemata and Northland for the North Island and in the Canterbury region for the South Island.

Figure 9 below presents the District Health Board (DHB) region that the Service's clients are associated with.

Figure 9: The Service's clients by DHB identification

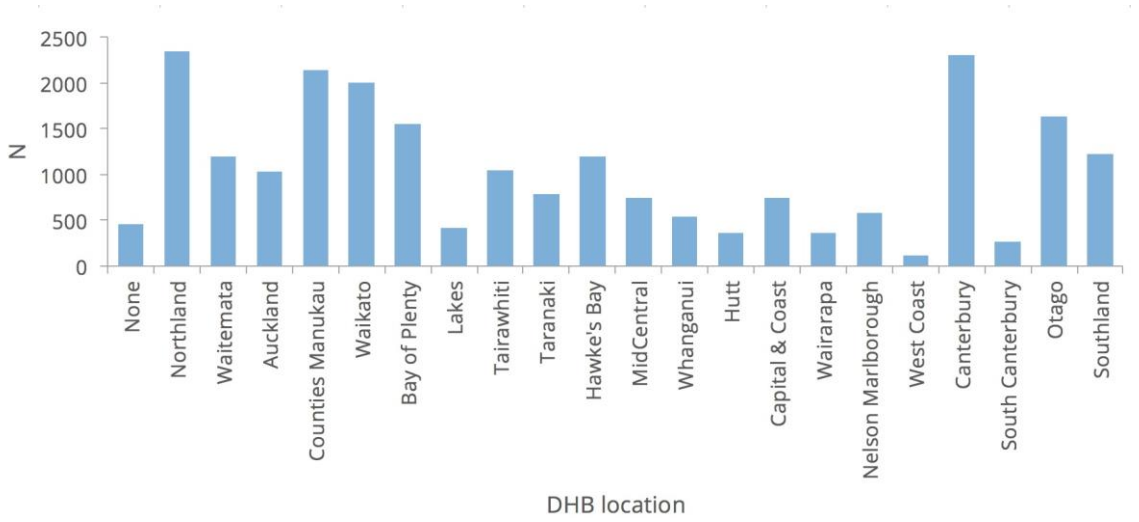


Figure 9 shows that the Service's intake is focused in Northland, Counties Manukau, Waikato and Bay of Plenty in the North Island and Canterbury in the South Island. This appears to fit reasonably closely the general profile of regions that have higher levels of abuse and/or deprivation as defined by MSD and Otago University respectively.

It should be noted that our analysis of the locational profile of the Service's clients has limitations. Using the two metrics of substantiated abuse and deprivation may not be the best indicators of general risk by region. There are also endogeneity concerns. For instance, Northland is one of the most deprived regions by the NZDep2013. However, there are relatively low substantiated findings of abuse in that region. This may be because abuse is under-reported in the region. However, as a general indication, it appears that the number of referrals to Stand by region coincides with the spread of the level of risk by region.

7.1.2 Clients' ages

Table 2 presents the distribution of ages for the Service's clients since the year 2000.

Table 2: Age of the Service's clients

Age at referral	% of total
Age < 5	< 1%
5	4%
6	8%
7	13%
8	16%
9	17%
10	17%
11	14%
12	10%
13	1%
13 < Age < 15	< 1%
Mean	9.0

Table 2 indicates that most of the children that are referred to the Service fall between the ages of seven and eleven, with almost all (approximately 99%) falling between the ages of five and thirteen. The average age at referral is nine years old.

Figure 10 shows the average age of the Service's clients for each year since 2000. While there has been some variation in the average age over time, the variation has been within a relatively narrow band of eight to nine years old.

Figure 10: Average age of the Service's clients by year

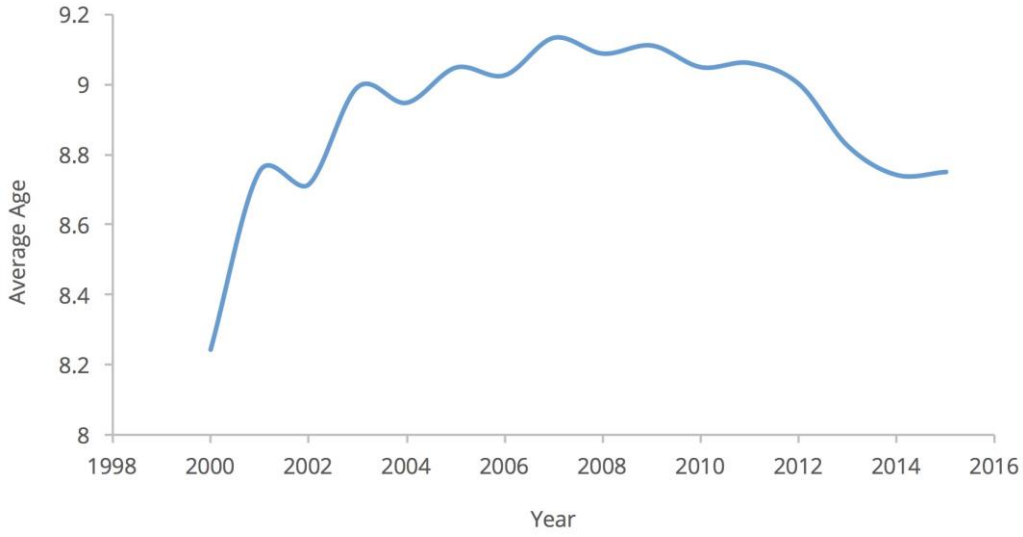


Figure 11 below presents the age frequency of the Service's clients by year. On the y-axis is the frequency of observations. On the x-axis is the age associated with that frequency and on the z-axis (the depth parameter) is the year.

Figure 11: Distribution of the Service's clients by age and year

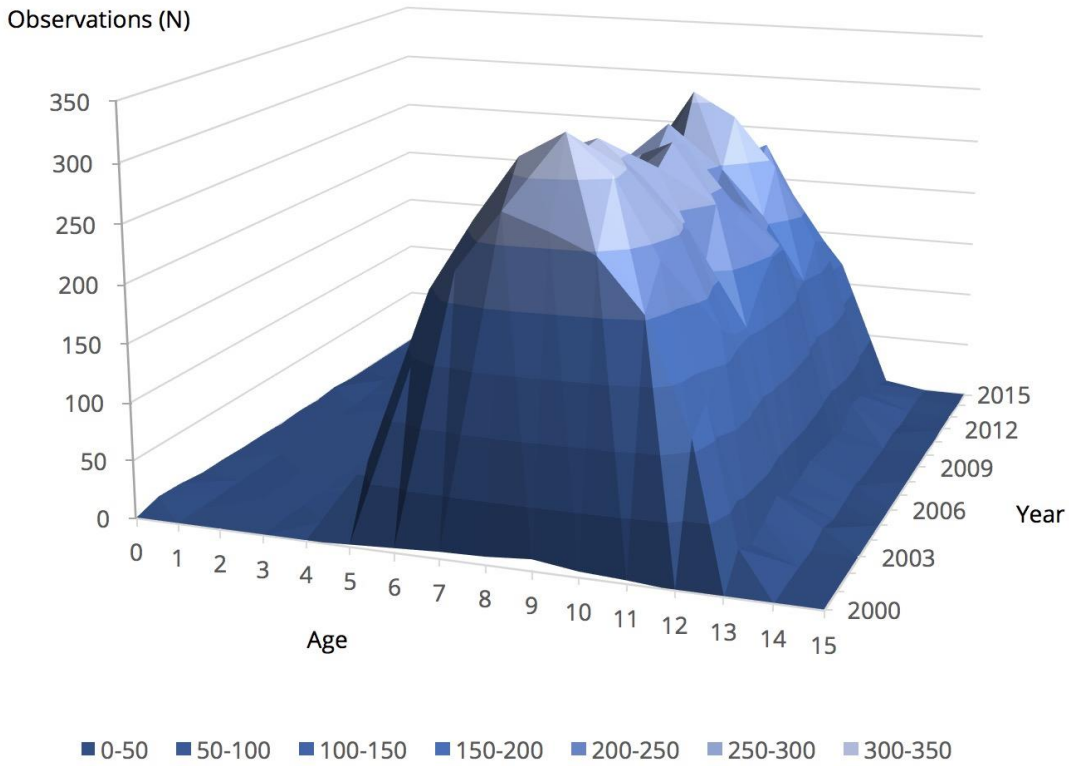


Figure 11 shows that most of the children in 2000 were between five and thirteen and that the highest frequency age was eight (with approximately 300 of the children being eight). Moving forward from 2000, we can see the distribution tightening around the seven to nine age brackets, with the peaks being around eight or nine years old.

7.1.3 Gender

Table 3 presents the gender ratio of the Service's clients since 2000. The table shows that males are over-represented in the Service's clients, accounting for 65% of clients on average.

Table 3: Gender of the Service's clients

Figure 12 below presents the gender ratio over time. It shows a small increase in the proportion of females (as of 2014, the distribution was 61% males and 39% females). Overall, a distribution of males outnumbering females by almost 2:1 is generally representative of the Service's client gender profile.

Figure 12: The Service's gender ratio over time

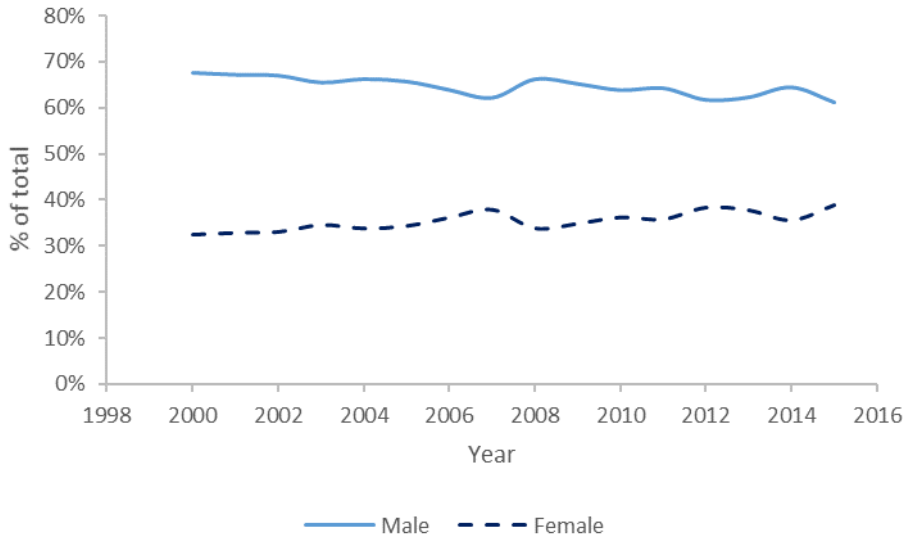


Figure 13 presents a time series of admission rates to CYF's Care & Protection and Youth Justice services. The grey represents Care & Protection (C&P) and the dark blue represents Youth Justice (YJ), with the solid lines showing the percentage of males and the dashed lines showing the percentage of females.

Figure 13: Admission rates to Care & Protection and Youth Justice services by gender

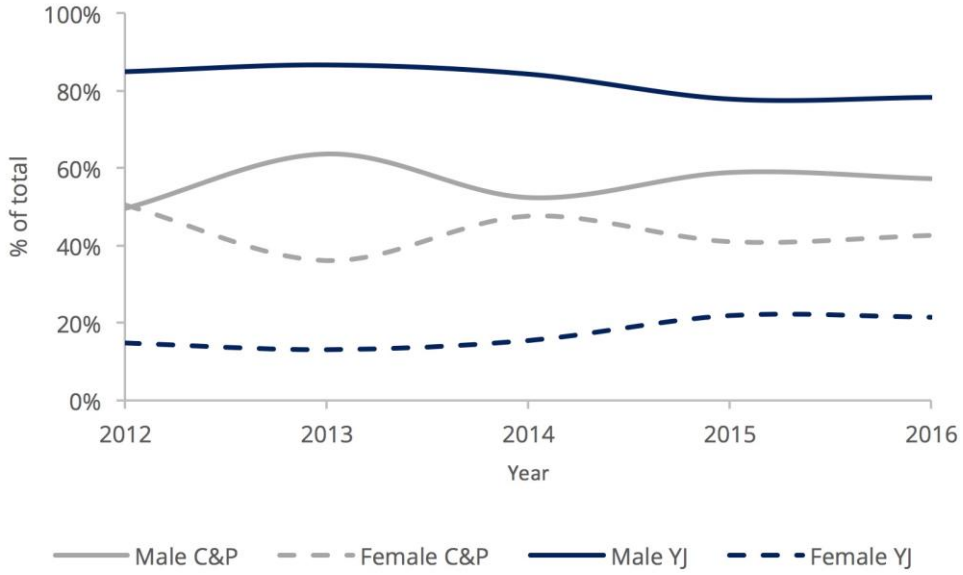


Figure 13 shows that since 2012, males have been on average slightly over-represented in the admissions to Care & Protection and heavily over-represented in the admissions to Youth Justice. This seems to generally indicate that the gender split of referrals to Stand roughly correlates with the gender split of at-risk children.

7.1.4 Ethnicity

Another demographic factor of interest is the ethnicity of the children referred to the Service. Table 4 presents the ethnicity percentage by major group since 2000.

Table 4: Ethnicity of the Service's clients since 2000

Table 4 shows that on average since 2000, 45% of the Service's clients were Maori, 46% were NZ European and 9% fell into the Other category.

Figure 14 breaks this down further by presenting the time series.

Figure 14: Ethnicity of the Service's clients by year

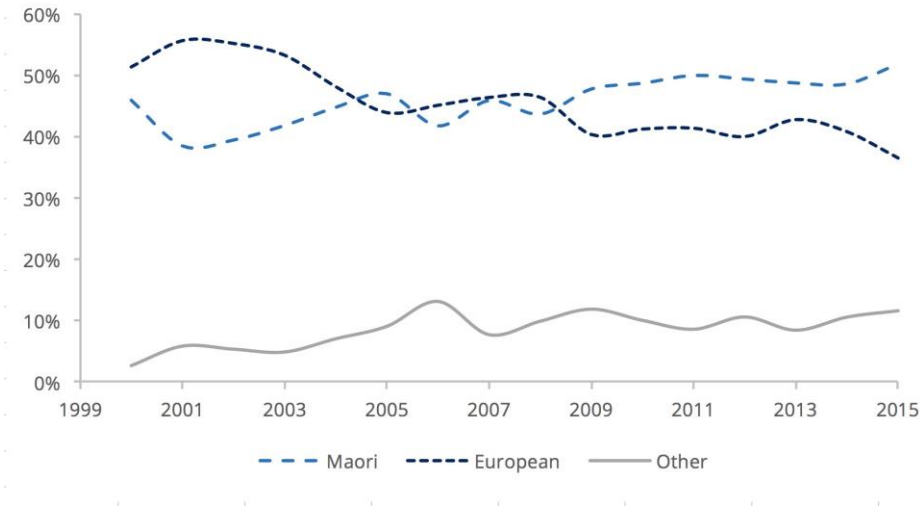


Figure 14 above shows that since the year 2000, the proportion of NZ European representation has been declining, whereas the proportions of Maori and Other have been increasing. In 2014, the proportions were 52% Maori, 37% NZ European and 12% other.

This trend is in the correct direction to be in-line with admittance rates to Care & Protection which are currently 64% Maori, 25% NZ European and 11% other. It is also moving to match current admittance rates to Youth Justice which are 71% Maori, 16% NZ European and 13% other. Furthermore, the NZ Treasury reports that of children who have on average one risk factor, 38% were NZ European and 40% were Maori. Of the children who had on average two risk factors, 26% were NZ European and 58% were Maori. This indicates that the ethnicity profile of the clients referred to the Service are consistent with on average "one to two risk-factor" children.

7.1.5 School decile

The last demographic metric considered is the school decile of the children served by Stand. This is depicted in Figure 15 below.

Figure 15: Distribution of the Service's clients' school decile since 2000

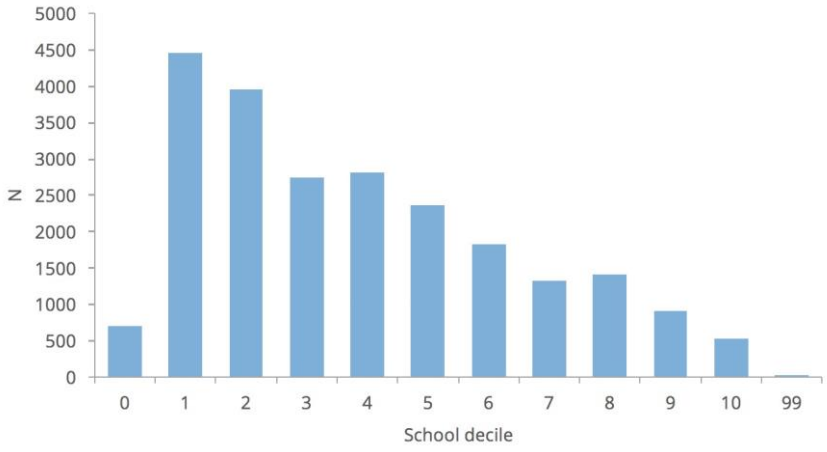


Figure 15 shows on average since 2000 the highest frequency for the decile of the schools the Service's clients come from is decile 1 schools. The frequency decreases

almost monotonically from decile 1 to decile 10. This shows that the distribution of the Service's clients is heavily skewed towards schools located in lower socio-economic areas. The deciles 0 and 99 in the figure above are unrated establishments (thought to be mostly home schools).

7.2 Client behavioural profiles

To further assess if the Service is reaching the children at higher risk of poor future outcomes, we investigate whether the behavioural profiles of children using the Service match the behavioural profiles we would expect from at-risk children, based on the evidence presented in the studies discussed in Sections 4 and 5.

7.2.1 Client average SDQ scores

Table 5 presents the average scores for the Service's clients before intervention for each of the constructs, beginning with the Total Difficulties as the first point of analysis, followed by the Impact score, the pro-social score and then the Total Difficulties constructs (being the four difficulties scores that make up the Total Difficulties score, emotional, conduct, hyperactivity, and peer problem scores).

Table 5: Average SDQ scores for full sample

	Total difficulties	Impact	Pro-Social	Emotion	Conduct	Hyper	Peer
Self	16.6	1.8	7.1	4.0	4.0	5.0	3.6
Parent	18.0	1.9	6.6	3.7	4.4	5.9	3.9
Teacher	15.9	2.1	5.4	3.0	3.4	5.7	3.8

Given the average scores in Table 5, Table 6 below applies the normal to abnormal rating system, defined in Section 5.1 Figure 2, to present at a high level the behavioural profile of the Service's clients.

Table 6: Classification of average SDQ scores

	Total difficulties	Impact	Pro-Social	Emotion	Conduct	Hyper	Peer
Self	Borderline	Borderline	Normal	Normal	Borderline	Borderline	Borderline
Parent	Abnormal	Borderline	Normal	Normal	Abnormal	Borderline	Borderline
Teacher	Borderline	Abnormal	Borderline	Normal	Borderline	Borderline	Borderline

Table 6 shows that, on average, the Service's clients have Total Difficulties and Impact scores classified as either borderline or abnormal. On average, the clients' pro-social scores are normal or borderline. In a general population sample, we would expect all constructs to be classified as normal on average, given the expected 80-10-10 distribution.

7.2.2 Caveats surrounding the data

There are important caveats surrounding the Stand SDQ data than should be noted.

Firstly, there is a concern around the self-assessment scores. Table 7 presents the sample sizes that are used to calculate each average SDQ score.

Table 7: Total observations used in each average SDQ score

	Total difficulties	Impact	Pro-Social	Emotion	Conduct	Hyper	Peer
Self	6,238	5,761	6,241	6,245	6,242	6,241	6,244
Parent	16,129	13,698	16,127	16,138	16,140	16,134	16,141
Teacher	16,344	15,710	16,344	16,351	16,351	16,350	16,351

In total, the Service has had 23,012 children referred to it between 2000 and 2015. Of these 23,012 referrals, 6,238 children filled out the pre-intervention self-evaluation SDQ correctly,[37] 16,129 children had credible Total Difficulties evaluations from their parent or caregiver and 16,344 had credible Total Difficulties evaluations from a teacher. The main reason for the much smaller sample size for the self-assessment than for the parent or teacher assessments is that only older children are thought to have the cognitive ability to assess themselves (the specific age where children assess themselves has appeared to have moved over time but is currently around eleven years old, as advised by Stand's staff).[38] Overall, we consider it is appropriate to place less weight on the self-assessments. This is because there are far less data points available and, depending on the state of the child's physiological well-being, the self-assessments that are undertaken may be more likely to be misleading.

In-line with the smaller sample size of the self-assessments, how comparative the three scores are, is called into question. The responses received from teachers and parents will cover all age groups referred to the Service, however the self-assessments will be eleven and older. For this reason, we want to be careful in comparing between the respondents and where possible analyse each separately. This is noted here but the three will be represented together for completeness.

Another important caveat to the sample size is that the difference between the (approximate) 23,000 children who have been through the programme and the (approximate) 16,000 for whom we have behavioural data becomes even larger when the analysis gets into post-intervention and the six-month follow up evaluations. This means that there is risk of endogeneity bias in the samples that are being analysed in this report. For example, it is possible that if a child appears to have extreme behavioural problems, then a reviewer may be more inclined to carry out their SDQ evaluation. Furthermore, on follow-up, a child who seems to have had a marked shift in behaviour may be more likely to have an SDQ evaluation based on the interest of the reviewer.

[37] The dataset received was large and as with any large dataset there is high risk of entry error, calculation of raw data error and formatting errors. The raw dataset was cleaned and cross checked thoroughly and it is our determination that there was no evidence of systematic errors that would undermine the analysis in general. For a summary of the data cleaning and checking process see Appendix 2.

[38] A breakdown of the SDQ's carried both by age and by year can be found in Appendix 3.

7.2.3 *Categorical distribution regarding the behavioural profiles of the Service's clients*

The second metric considered is the distribution of normal, borderline and abnormal scores. As mentioned, the SDQ scoring system was designed to have 80% of respondents from the general population being classified as having a normal score, 10% a borderline score and 10% an abnormal score, based on a large U.K. cohort study and confirmed by a NZ based study.

Table 8 below shows the proportions of normal, borderline and abnormal SDQ scores of children who use Stand, before intervention.

Table 8: Distribution of full sample classifications for Total Difficulties SDQ

	Normal	Borderline	Abnormal	N
Self	43%	25%	32%	6,238
Parent	26%	15%	60%	16,129
Teacher	28%	18%	54%	16,344

The distributions outlined in Table 8 for the Service's clients are quite different from what would be expected in a random sample from New Zealand general population. If the UK metrics can be reliably transferred to New Zealand, we would expect, for the general population, to see 80% normal, 10% borderline and 10% abnormal scores. In contrast, Table 8 shows that for the Service's clients, of the self SDQ response scores, 43% are classified as normal, 25% as borderline and 32% as normal. The parent-responses classify 26% of the children's scores as normal, 15% as borderline and 60% as abnormal. The teacher-responses classify 28% of the children's scores as normal, 18% as borderline and 54% as abnormal. A full breakdown of the distributions for all SDQ constructs can be found in Appendix 4.

7.2.4 *A comparison of risk-behaviour using the Growing Up in New Zealand study*

Drawing again from Figure 4 in Section 5.2, the Growing Up in New Zealand study classified its cohort based on a set of environmental factors before the child was born. After two years, the parents of the cohort evaluated the behaviour of the child using the SDQ Total Difficulties constructs. Table 9 below again represents the behavioural distributions from normal to abnormal by classified environmental risk group.[39]

[39] While not seemingly specifically noted in the study it is likely that the Study utilised the parent SDQ as the children were two years old.

Table 9: GUiNZ-SDQ risk profile

GUiNZ-SDQ risk profile

		Behavioural Risk (SDQ Classification)			
		Normal	Borderline	Abormal	N
Enviormental risk (GUiNZ)	Low risk	79%	12%	8%	2,859
	Medium risk	63%	18%	19%	2,715
	High risk	37%	19%	44%	739

Table 9 depicts an indication that increasing environmental risk is associated with increased behavioural risk for a cohort of NZ two year olds. That is, in the low environmental risk category defined by GUiNZ, 8% are classified as having abnormal total behavioural difficulties. In the medium environmental risk category, this increases to 19% being classified as having abnormal total behavioural difficulties. Lastly, in the high environmental risk category, 44% are classified as having abnormal total behavioural difficulties. This indicates a strong correlation between the environment a child is born into and behavioural outcomes at the age of two.

7.2.5 A simple weighted index as a tool for comparison

One simple way to view the distributions of SDQ normal to abnormal classifications is to take a weighted average of the percentages that fall into each of the SDQ classifications. The equation below presents a way to normalise the distribution of normal to abnormal classifications into a single metric. We have called this the "Behaviour profile score" and it is calculated by summing the percent of the sample of SDQ's classified as normal, the percentage of the sample that were classified as borderline multiplied by two, and the percentage of the sample that were classified as abnormal multiplied by three. [40]

Equation 1: Behaviour profile score

$$Behaviour\ profile\ score = Normal\ \% * 1 + Borderline\ \% * 2 + Abnormal\ \% * 3$$

The number that is produced by Equation 1 will be between one and three.[41] A behaviour profile score of exactly one would constitute a complete sample of children classified as normal. A behaviour profile score of exactly three would constitute a complete sample of children classified as abnormal.

Table 10 presents the behavioural profile scores for the three environmental risk groups defined by GUiNZ.

Table 10: GUiNZ behaviour profile scores by environmental risk

Low risk	Medium risk	High risk
1.29	1.56	2.07

[40] We are unable to compare average SDQ scores due to a due to not having the GUiNZ raw data, and GUiNZ not reporting them.
[41] Proporrtions are used for the percentages, ie 0.19=91%

Table 10 shows that the low-risk cohort has a behaviour profile score of 1.29. The medium-risk cohort has a behaviour profile score of 1.56 and the high-risk cohort has a behaviour profile score of 2.07.

Table 11 below presents the behaviour profile scores for the observed distribution of Stand's clients, by respondent.

Table 11: Stand Total Difficulties behaviour profile scores

Self	Parent	Teacher
1.89	2.34	2.26

Table 11 shows that the self SDQ gives a behaviour profile score of 1.89. The parent SDQ gives a behaviour profile score of 2.34 and the teacher SDQ gives a behaviour profile score of 2.26.

Figure 16 below presents the implications of Table 10 and 11 graphically. The vertical lines are the GUiNZ behaviour profile scores, by environmental risk group. The plotted points are Stand's behaviour profile scores, by respondent. The x-axis and y-axis are both behaviour profile scores that go from one to three and give the 45-degree line of fit that the points can move along.

Figure 16: Stand Total Difficulties behaviour profile scores

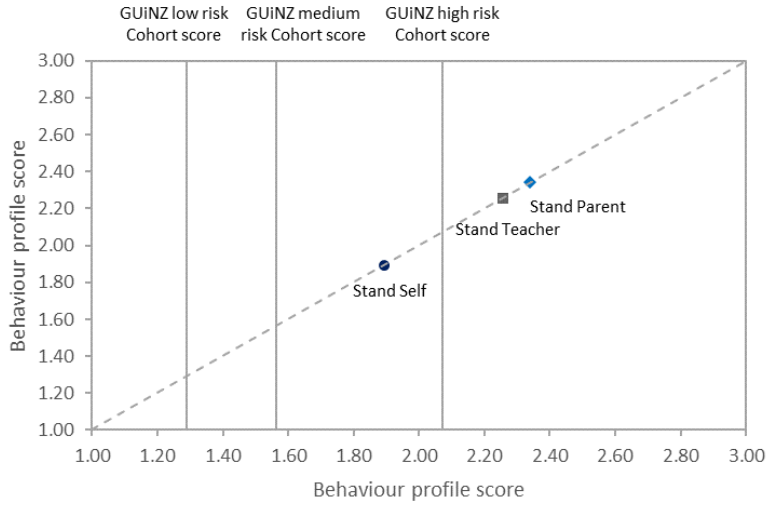


Figure 16 above shows that if the GUiNZ distributions are comparable to that of Stand's clients in any form, then the self-reported SDQ distribution for Stand's clients shows a higher behaviour profile score than the medium-risk GUiNZ cohort. Also, Stand's parent and teacher-reported SDQ distribution show a higher behaviour profile score than the high-risk GUiNZ cohort.

In conducting this analysis, again we note that the GUiNZ cohort was assessed at two-years old, however the data from Stand's service, is SDQ assessments made of children between 0-14 years old.[42] This likely creates many problems with the comparison between the GUiNZ data and the data we attained from Stand. However, in the absence of further data of more comparable NZ based SDQ assessments and to keep sample size

[42] See Appendix 3 for more details.

as high as possible we will continue with this analysis while noting that further work should be done in this respect.

7.3 Summary of Stand's client profiles

This section has aimed to address the question 'is Stand targeting the right children?' Overall, we consider Stand successful in targeting children who are at higher risk than normal.

We have seen that demographically:

- it appears that the Service has greatest demand/highest activity levels on areas of the country that are higher risk, measured by substantiated finding of abuse or have higher general deprivation;
- the Service is referred clients who are on average approximately nine years old when they are referred;
- clients are on average approximately 50:50 NZ European and Maori clients with an increasing trend of Maori and other ethnicities and a decreasing trend of NZ European clients through time, which is converging to what is seen in the Youth Justice and the Care & Protection admittance data;
- the Service is referred almost 2:1 males to females, consistent with what might be expected with admissions to Youth Justice and Care & Protection services; and
- the Service is referred clients who are far more likely to come from low deciles schools.

Behaviourally, we have seen that Stand's clients:

- on average, exhibit borderline to abnormal total behavioural difficulties;
- appear to have behavioural distributions that are more skewed toward risk than expected for the general New Zealand population; and
- on average, exhibit a behavioural profile with greater Total Difficulties scores than the 'high environmental risk' cohort in the Growing Up in New Zealand study.

8. The impact of the Stand for Children Service on behaviour profiles

This section seeks to address the second key question of the Stand for Children Service:

Is Stand having a positive impact on the children it serves?

We use the same SDQ behavioural indicators presented in the previous section to analyse the effect Stand has on its clients from pre-intervention to post-intervention and a six-month follow-up. Section 8.1 breaks down the average behavioural scores for each of the SDQ constructs. Section 8.2 then looks at the distribution of the behavioural profile and how that relates to the GUiNZ environmental risk group. Lastly, Section 8.3 we analyse the proportion of clients that experienced positive and negative behavioural changes and the size and consistency through time of those changes. Section 8.4 provides a summary.

8.1 Average scores

Table 12 below outlines the average SDQ scores for the full sample of Stand's clients at three stages: pre-intervention; post-intervention; and six-month follow-up.

Table 12: Average SDQ scores through the intervention process

Pre-intervention

	Total difficulties	Impact	Pro-Social	Emotion	Conduct	Hyper	Peer
Self	16.6	1.8	7.1	4.0	4.0	5.0	3.6
Parent	18.0	1.9	6.6	3.7	4.4	5.9	3.9
Teacher	15.9	2.1	5.4	3.0	3.4	5.7	3.8

Post-intervention

	Total difficulties	Impact	Pro-Social	Emotion	Conduct	Hyper	Peer
Self	14.5	1.4	7.2	3.4	3.3	4.5	3.2
Parent	14.5	1.3	7.0	2.8	3.4	4.9	3.4
Teacher	14.0	1.3	5.6	2.7	2.9	4.9	3.6

6m follow-up

	Total difficulties	Impact	Pro-Social	Emotion	Conduct	Hyper	Peer
Self	14.9	2.2	7.1	3.7	3.3	4.5	3.4
Parent	14.8	1.6	6.9	3.0	3.6	5.0	3.3
Teacher	13.6	1.5	5.7	2.5	2.7	5.0	3.4

Table 13 depicts the change in each of the average scores. The top section of the table shows the absolute change for each average score from pre-intervention to post-

intervention and the lower section of the table depicts the absolute change from pre-intervention to the six-month follow-up evaluation.

Table 13: Absolute changes in SDQ scores

Δ Pre - post intervention

	Total difficulties	Impact	Pro-Social	Emotion	Conduct	Hyper	Peer
Self	-2.1**	-0.4**	0.1**	-0.6**	-0.7**	-0.5**	-0.4**
Parent	-3.5**	-0.6**	0.4**	-1.0**	-1.0**	-0.9**	-0.6**
Teacher	-2.0**	-0.8**	0.3**	-0.3**	-0.6**	-0.8**	-0.3**

Δ Pre - 6m follow-up

	Total difficulties	Impact	Pro-Social	Emotion	Conduct	Hyper	Peer
Self	-1.7**	0.4**	0.0	-0.3**	-0.7**	-0.5**	-0.2**
Parent	-3.2**	-0.3**	0.3**	-0.8**	-0.9**	-0.9**	-0.6**
Teacher	-2.4**	-0.6**	0.3**	-0.5**	-0.7**	-0.7**	-0.5**

** statistically different from 0 at p < 0.01 level; * statistically different from 0 at p < 0.05 level.[43]

The top section of Table 13 shows that from pre-intervention to post-intervention, each respondent group (self, parent and teacher) on average reported a decrease in the children's Total Difficulties score, Impact score and in all four of the Total Difficulties constructs. Also, each the respondents have reported an increase in the pro-social scores. This indicates that there have been, on average, behavioural improvements from pre-intervention to post-intervention across every metric.

The lower section of Table 13 shows the changes from pre-intervention to the six-month follow-up. There is still a decrease (meaning a positive behavioural change) in the average Total Difficulties scores and each of its constructs. There is a slight reversion for the average self and parent Total Difficulties scores (relative to the post-intervention) and a continued improvement on average for the teacher-assessed Total Difficulties score. Also, there is an overall increase (or negative behavioural shift) in the self-reported Impact score. All the self-reported six-month follow-up metrics, however, should be treated with caution as they are based on a sample of only 354, as is outlined in Table 15 below, and discussed subsequently.

Table 14 presents the pre-intervention, post-intervention and the follow-up classifications of behaviour implied by the average scores using the SDQ Total Difficulties scoring process.

[43] Significance calculated as: $Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$; where $\mu_1 - \mu_2$ is the null hypothesis and assumed to be zero. Tests have been conducted with critical values consistent with two tailed tests which may be best in this setting but serves as a high-level indicator.

Table 14: SDQ behaviour classification through the intervention process

Pre-intervention

	Total difficulties	Impact	Pro-Social	Emotion	Conduct	Hyper	Peer
Self	Borderline	Borderline	Normal	Normal	Borderline	Borderline	Borderline
Parent	Abnormal	Borderline	Normal	Normal	Abnormal	Borderline	Borderline
Teacher	Borderline	Abnormal	Borderline	Normal	Borderline	Borderline	Borderline

Post-intervention

	Total difficulties	Impact	Pro-Social	Emotion	Conduct	Hyper	Peer
Self	Normal	Borderline	Normal	Normal	Borderline	Normal	Borderline
Parent	Borderline	Borderline	Normal	Normal	Borderline	Normal	Borderline
Teacher	Borderline	Borderline	Borderline	Normal	Borderline	Normal	Borderline

6m follow-up

	Total difficulties	Impact	Pro-Social	Emotion	Conduct	Hyper	Peer
Self	Normal	Abnormal	Normal	Normal	Borderline	Normal	Borderline
Parent	Borderline	Borderline	Normal	Normal	Borderline	Normal	Borderline
Teacher	Borderline	Borderline	Borderline	Normal	Borderline	Normal	Borderline

Table 14 above shows that from pre-intervention to post-intervention, the self-reported average Total Difficulties classification moves from a borderline to a normal. This classification change is maintained at the six-month follow-up. The parent-reported Total Difficulties classification also changes from abnormal to borderline and is also maintained at follow-up. The teacher-reported average Total Difficulties classification does not change but the Impact classification changes from abnormal to borderline and the change does not revert at six-month follow-up. The self-reported Impact score on the other hand does not change classification group from pre to post but changes from borderline to abnormal (on average) from post to follow-up. It is also possible that after the intervention the child is becoming more conscious of their surroundings and can more acutely assess the risk they are subject to.

For the individual Total Difficulties constructs, the hyperactive score shifts from borderline to normal for all respondents and the parent conduct score shifts from abnormal to borderline, all of which do not revert at follow-up.

It is important to note that these are high-level classifications that have been developed from observed data. While it is unclear exactly what a sustained shift of behavioural classification may mean, the change would appear to be relevant considering there are only three classification groups.

Table 15 shows the sample sizes used in the preceding analysis.

Table 15: Observations used in calculating average SDQ scores

Pre-intervention

	Total difficulties	Impact	Pro-Social	Emotion	Conduct	Hyper	Peer
Self	6,238	5,761	6,241	6,245	6,242	6,241	6,244
Parent	16,129	13,698	16,127	16,138	16,140	16,134	16,141
Teacher	16,344	15,710	16,344	16,351	16,351	16,350	16,351

Post-intervention

	Total difficulties	Impact	Pro-Social	Emotion	Conduct	Hyper	Peer
Self	2,378	2,352	2,370	2,378	2,378	2,378	2,378
Parent	7,871	7,806	7,861	7,872	7,874	7,873	7,874
Teacher	7,944	7,854	7,936	7,947	7,948	7,947	7,946

6m follow-up

	Total difficulties	Impact	Pro-Social	Emotion	Conduct	Hyper	Peer
Self	362	354	360	362	362	362	362
Parent	1,397	1,377	1,396	1,399	1,398	1,398	1,399
Teacher	2,584	2,555	2,582	2,584	2,584	2,584	2,584

Table 15 above raises important caveats further to the endogeneity risk to the analysis undertaken. As in the previous section, the dataset contained 6,238 observed and credible self-reported Total Difficulties scores, 16,129 observed and credible parent-reported Total Difficulties scores and 16,344 observed and credible teacher-reported Total Difficulties scores. At post-intervention, these numbers dropped to 2,378 self-reported, 7,871 parent-reported and 7,944 teacher-reported. At a six-month follow-up evaluation, these numbers dropped further to 362 self-reported, 1,397 parent-reported and 2,584 teacher-reported. Because the average scores are being used to map changes in the behavioural profile of the children, there is a risk of endogeneity within the sampling process (as discussed in Section 7.2 above). There is also a risk of omitted variable bias because we are retaining the full sample at each point to maximise the size of each sample. Also, the extreme decline in the number of observations for the six-month follow-up SDQ reports. This likely makes the six-month records less reliable.

The risk of omitted variable bias from different sample sizes from pre to post to six-month follow-up has been tested for by using a cohort approach in Appendix 5. That is, the above analysis has been reconstructed using firstly only those children who have all three observations, secondly testing the average absolute change in each score of only the children who have both pre and post intervention observations and lastly testing the change of only those children who have pre and six-month follow-up observations. As shown in Appendix 5 there is no material change to the findings when any of the different methodologies are adopted. This is a useful cross-check but, the endogeneity risk remains.

8.2 Change of distributions for Total Difficulties scores

As in the previous section, the second metric that is considered is the distribution of the normal to abnormal classifications for the Total Difficulties scores.[44] Table 16 presents the distribution of each sample period with firstly pre-intervention, then post intervention and lastly follow-up. The first column is the respondent. The second column presents the percentage of the total responses received from the respondent for that evaluation that were classified as normal. The third is the percentage classified as borderline, the fourth is the percentage classified as abnormal and the last is the sample size (N).

Table 16: Distribution of SDQ behaviour profile through intervention

Pre-intervention				
	Normal	Borderline	Abnormal	N
Self	43%	25%	32%	6,238
Parent	26%	15%	60%	16,129
Teacher	28%	18%	54%	16,344
Post-intervention				
	Normal	Borderline	Abnormal	N
Self	57%	21%	21%	2,378
Parent	47%	15%	38%	7,871
Teacher	39%	19%	42%	7,944
6m follow-up				
	Normal	Borderline	Abnormal	N
Self	55%	17%	29%	362
Parent	46%	13%	41%	1,397
Teacher	42%	17%	41%	2,584

Table 16 shows the presents the distributions of the normal to abnormal SDQ classifications for the clients of the Service firstly pre-intervention, followed by post-intervention and lastly at the six-month follow up assessment.

Table 17 below shows the change from pre to post (in the top section of the table) and from pre to follow-up (in the lower section of the table).

[44] Again, Total Difficulties are applied here because they are the most widely accepted measure but a breakdown of the individual constructs can be found in Appendix 6.

Table 17: Change in distribution of SDQ behaviour profile through intervention

Δ Pre - post intervention			
	Normal	Borderline	Abnormal
Self	15%	-4%	-10%
Parent	22%	0%	-22%
Teacher	11%	1%	-12%
Δ Pre - 6m follow-up			
	Normal	Borderline	Abnormal
Self	12%	-9%	-3%
Parent	21%	-2%	-19%
Teacher	13%	0%	-13%

Table 17 above shows a consistent decrease in the percentage of children that were classified as abnormal and a consistent increase in the percentage of children that were classified as normal from pre to post-intervention and from pre-intervention to follow-up.

8.2.1 GUiNZ and Stand's behaviour profile score through intervention

Consistent with the previous section on defining the baseline of Stand's client behavioural profiles, as a high-level indication of what these changes in behavioural distribution may mean the behaviour profile score is again adopted as specified in Equation 1.

Table 18 below presents the calculated behaviour profile scores for the three respondent groups at the three different stages as carried out by Stand.

Table 18: Stand behaviour profile scores throughout intervention

	Self	Parent	Teacher
Pre	1.89	2.34	2.26
Post	1.64	1.91	2.03
6m	1.74	1.95	1.99

Figure 17 below graphically depicts the pre, post and follow-up behaviour profile scores for the self-assessment.

Figure 17: Stand self-responses vs. GUiNZ behaviour profile scores

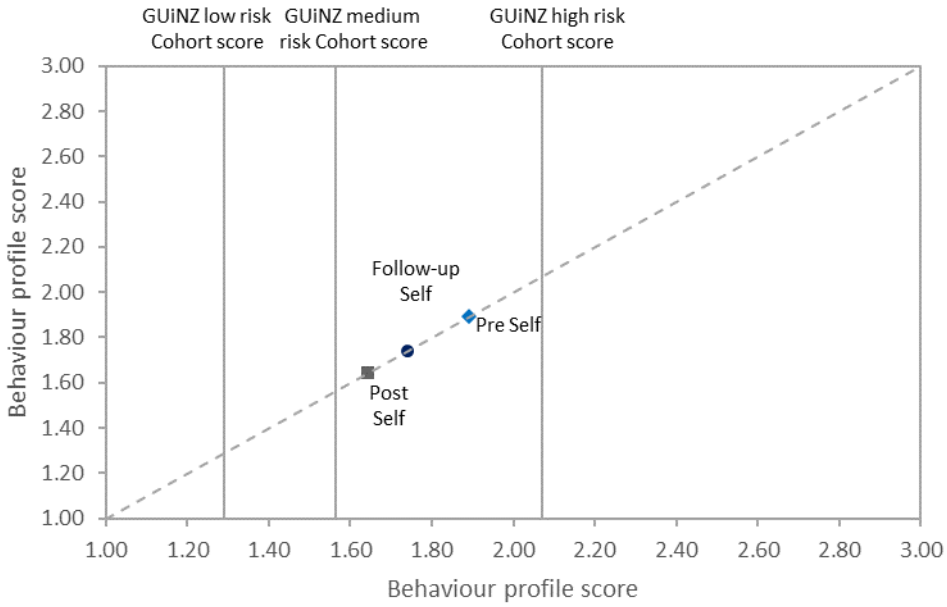


Figure 17 shows that there is a decrease in the profile score from pre to post-intervention that remains below the GUiNZ high environmental risk group behaviour score and above the medium environmental risk group behaviour score. At follow-up, this score had increased back towards the pre-intervention but by a smaller magnitude.

As previously noted, we want to be careful in not overstating the comparison between the Service's clients and the GUiNZ findings, particularly for the self-assessments. The age of the self-assessments will be vastly different from the sample of two-year-olds given by the GUiNZ cohort. The self-assessment comparison has been analysed here for completeness of the analysis (being that regardless of what the changes mean in relation to GUiNZ there is an observed improvement in the behaviour profile on average).

Figure 18 below depicts the change in the behaviour profile score for parent-responses from pre to post-intervention, to six-month follow-up.

Figure 18: Stand self-responses vs. GUiNZ behaviour profile scores

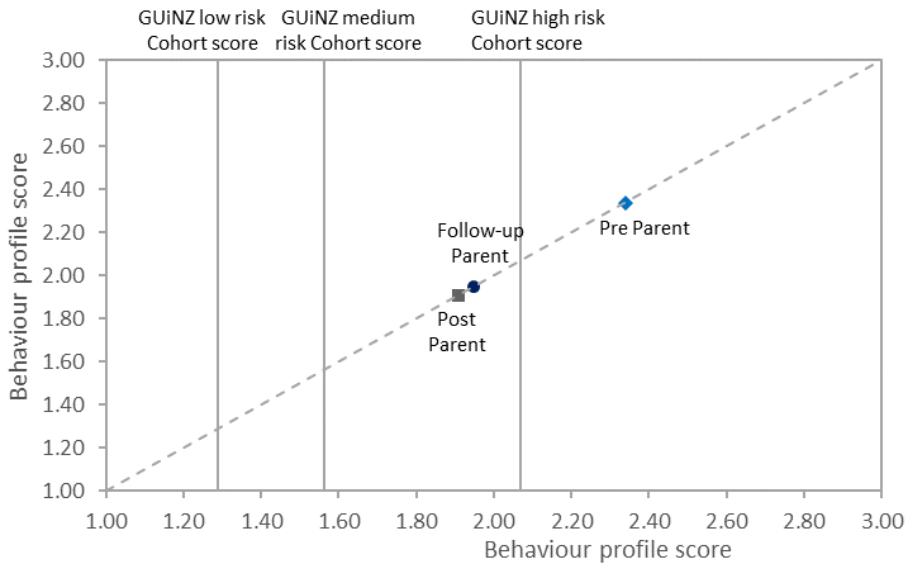


Figure 18 shows quite a dramatic decrease in the profile score from pre to post-intervention that moves from above the GUiNZ high environmental risk group behaviour score to below it. At follow-up, this score had a small increase back towards the pre-intervention but has remained below the GUiNZ high environmental risk group behaviour score.

Figure 19 below depicts the change in the behaviour profile score for the teacher-responses from pre to post-intervention, to six-month follow-up.

Figure 19: Stand self-responses vs. GUiNZ behaviour profile scores

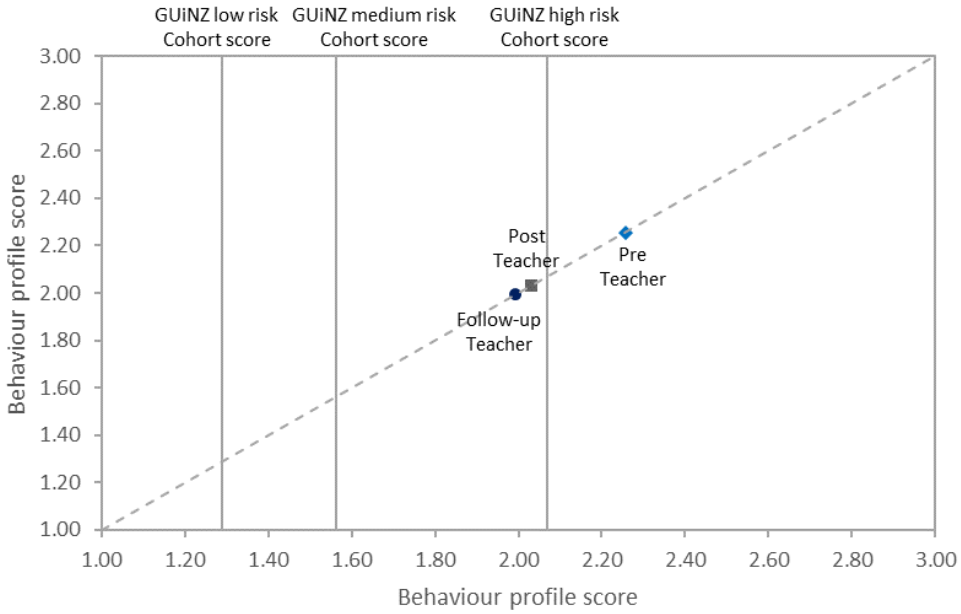


Figure 19 shows a smaller decrease (relative to the parent-responses) in the profile score from pre to post-intervention that moves from above the GUiNZ high environmental risk group behaviour score to below it. At follow-up, this score had continued to decrease, albeit at a slower rate, from the pre-intervention profile score.

8.3 Improvement proportions and magnitudes

The final consideration is to break down the changes in behaviour to look more specifically at the changes and the magnitudes as percentages. Table 19 below presents the number counts of increased SDQ Total Difficulties scores from pre to post-intervention and from pre to follow-up. It breaks the numbers down by the number that have experienced decreased scores (meaning a positive behavioural impact), the number that have experienced no change and the number that have experienced increased scores (negative behavioural impact) and then the total observations in each sample (for each evaluation and each respondent).

Consistent with our prior analysis, we have adopted the Total Difficulties scores as they are the most widely used. However, Appendix 6 presents a breakdown of each of the SDQ constructs.

Table 19: Total Difficulties score observed directional shifts

Relative to pre-intervention	N Decreased	N No change	N Increased	N	% Decreased	% No Change	% Increased
Post Intervention SDQ Total Difficulties							
Self	1,166	152	680	1,998	58%	8%	34%
Parent	4,822	509	1,971	7,302	66%	7%	27%
Teacher	4,158	511	2,787	7,456	56%	7%	37%
Six Month Follow Up SDQ Total Difficulties							
Self	155	27	122	304	51%	9%	40%
Parent	781	76	413	1,270	61%	6%	33%
Teacher	1,323	148	904	2,375	56%	6%	38%

Table 19 above shows that of the self-responses, 58% had experienced a positive behavioural shift, 8% had effectively not changed and 34% had experienced a negative behavioural change. On follow-up, these numbers changed to 51% had experienced a positive behavioural shift, 9% had effectively not changed and 40% had experienced a negative behavioural change.

Of the parent-responses, 66% had experienced a positive behavioural shift, 7% had effectively not changed and 27% had experienced a negative behavioural change. On follow-up, these numbers changed to 61% had experienced a positive behavioural shift, 6% had effectively not changed and 33% had experienced a negative behavioural change.

Of the teacher-responses, 56% had experienced a positive behavioural shift, 7% had effectively not changed and 37% had experienced a negative behavioural change. On follow-up, these numbers changed to 56% had experienced a positive behavioural shift, 6% had effectively not changed and 38% had experienced a negative behavioural change.

Overall there is a positive improvement for a majority of the children but there is negative change for a sizeable minority and a small number of children have experienced no change.

Table 20 below presents the average percentage change in Total Difficulties score for each of the sub-samples and the full sample, by respondent and by evaluation.

Table 20: Total Difficulties score observed average percentage changes

Relative to pre-intervention	Mean decreased	Mean increased	Mean full sample
Post intervention SDQ Total Difficulties			
Self	-33%**	49%**	-2%
Parent	-36%**	45%**	-11%**
Teacher	-38%**	89%**	11%**
Six Month Follow Up SDQ Total Difficulties			
Self	-35%**	81%**	14%
Parent	-37%**	44%**	-8%**
Teacher	-41%**	98%**	14%**

** statistically different from 0 at p < 0.01 level; * statistically different from 0 at p < 0.05 level

Table 20 shows that of the self-responses positively affected (Mean decreased), the behavioural profile was reduced by approximately a third. On follow-up, this decrease was slightly larger, which could indicate that of those positively affected by the treatment the affect is on-going. Of the parent and teacher-responses, the average change for those positively affected was more than a third and the numbers exhibit the same increasing trend on follow-up. One important note is that the averages of the full sample may be misleading in this setting where the constructs are absolute. If the profile was to be doubled then the result would be an increase of 100%. On the other side, if the profile was halved (arguably an equal and opposite effect) the change would be only -50%. This demonstrates the unitary problems encountered. Looking at the changes in terms of natural logs may help resolve this. However, for the time being we focus on the means of the increased vs. decreased samples.

8.4 Summary of the Service's impact

This section has aimed at addressing the question: "Does Stand make an impact?" We have found that:

- for the average SDQ scores, there is change in a positive direction on almost every behavioural metric from pre to post-intervention to follow-up, there is a consistent indication across the three assessors (self, parent and teacher) that the SDQ classifications and distributions change for the better on average;
- in the parent and teacher assessments of behaviour there is indication that on average the children referred to the Service have a change in behaviour profile that shifts from being above the behaviour profile of the high environmental risk group to below it (effectively moving the children from the profile consistent with a high-risk environment to a medium-risk environment profile);
- more than half show positive behavioural shifts after the intervention and after six-months of on average a 1/3 improvement in their behavioural profile; and
- the Service does not have a positive effect on all its referrals with around 30% recording an adverse change.

9. Fiscal saving and economic benefit estimates for the Stand for Children Service

This section considers the final key question of the Stand for Children Service:

What is the long-term fiscal impact of Stand?

To address this question, it must be noted that the available data and the current position of the literature limits significantly what can be said. From our earlier analysis, it appears that Stand is targeting, on average, children in the correct risk range and it also appears that Stand has a positive impact, on average, on the behaviour profiles of children it serves. However, the available data at this stage does not permit a robust connection to be drawn between small and medium changes in behaviour, as measured by SDQ scores, and long-term fiscal impacts.

Given the data limitations, the most feasible approach to address the question of the fiscal impact is to "reverse engineer" the question by asking, 'what is the minimum effect the Service would need to have to produce long-term fiscal benefits for the government?' This section presents that analysis.

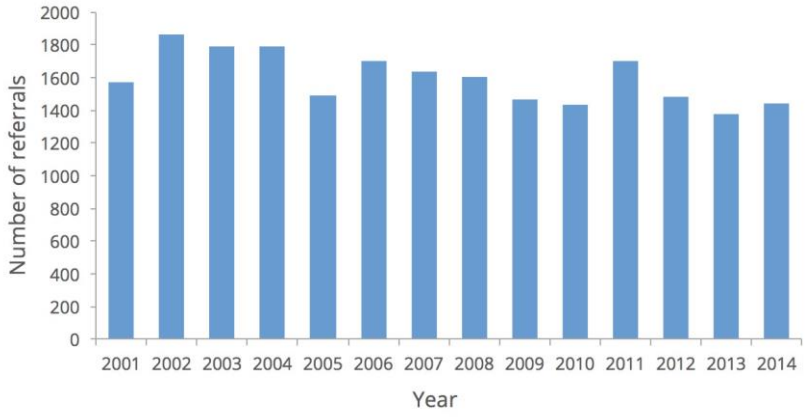
9.1 Producing a long-term positive financial outcome for the government

9.1.1 The Stand for Children Service's funding and referral numbers

Stand receives $16,011,835 in government funding per year, primarily from the Ministry of Social Development (MSD), and also from the Ministry of Education (MoE).

Turning to numbers of referrals, Figure 20 below presents the number of referrals accepted by Stand between 2001 and 2014.[45]

Figure 20: Stand's number of accepted referrals per year



[45] 2000 and 2015 have not been included because they were incomplete years in the dataset we received from Stand.

Figure 20 shows a relatively stable number of referrals between 2001 and 2014. The average number of referrals is 1,596 per year.

Combining the cost to the government of the Service with the average number of referrals indicates that Stand receives effectively $10,030 in government funding per child.

9.1.2 Recap on the NZ Treasury's cost estimates

Table 21 below re-presents the Treasury's estimates of the cost to the government for children with on average one, two, three or four risk factors.[46,47]

Table 21: Recap of risk factors and NZ Treasury estimates of direct cost to NZ

Risk factors	N	Treasury estimate of average total cost before 21	Treasury estimate of average total cost before 35
Total sample	510,351	$29,400	$67,700
0	354,864	$10,800	$33,300
1	87,234	$47,100	$100,300
2	44,142	$90,500	$180,300
3	19,857	$123,300	$244,900
4	4,254	$139,300	$275,300
2++	68,250	$103,100	$205,000
3++	24,111	$126,100	$250,300

9.1.3 Fiscal effect of decreasing by one Treasury risk factor

Table 22 below presents the estimated fiscal cost saving from changing a child's rating by one risk factor.[48] Up to this point we have noted that Stand, like other similar interventions, is unable to directly impact on the specific environmental risk factors identified in the Treasury report (such as the child's mother's formal qualification or a parent's past prison conviction). However, to assign a fiscal value to the intervention we assume for the time that the risk factors are figurative. That is, if the Service can have a positive and lasting effect on the child and family of an at-risk child it will still decrease the risk of poor future life outcomes (despite not being able to address the specific environmental factors directly).

[46] "Characteristics of Children at Greater Risk of Poor Outcomes as Adults", New Zealand Treasury Analytics and Insights Team, February 2016.
[47] 2014 real NZD terms.
[48] As previously noted, as per the Treasury's February report, these estimates are in 2014 dollar terms and are not discounted into present value terms. We don't know the timing of the cash flows and presumably the timing would change for each child so obtaining robust present values is unfeasible in this setting. Therefore, we take Treasury's estimates at face value but note they are not present values.

Table 22: Implied cost saving by shifting a child the equivalent of one NZ Treasury risk factor

Movement of one risk factor	Treasury estimate of average total cost before 21	Treasury estimate of average total cost before 35
Stand takes a 1 to a 0	$36,300	$67,000
Stand takes a 2 to a 1	$43,400	$80,000
Stand takes a 3 to a 2	$32,800	$64,600
Stand takes a 4 to a 3	$16,000	$30,400
Average saving per child	**$32,125**	**$60,500**

Table 22 shows that if the Service changes a child's risk profile by the equivalent of one Treasury risk factor, this can be expected to save the government $32,125 by the age of 21, increasing to $60,500 by the age of 35 per child.

Given the average cost per child to the government of the Service is $10,030 and the return to the government from shifting a child by one risk factor (a "successful outcome") is $60,500 – based on the costs savings to age 35 – the return to the government on a "successful outcome" $\frac{\$60,500}{10,030} = 6$. Or to express the finding another way, for the government to break-even over time in its funding of the Service, Stand must have a lasting positive effect that is equivalent to shifting a child by one Treasury risk factor for one child in every six referred to the Service.[49]

9.1.4 Fiscal effect of decreasing by two Treasury risk factors

Table 23 below shows the average financial savings to the government if Stand changes a child's risk profile by effectively two Treasury risk factors. By the age of 21 this would equate to an average saving of $68,233 and by age 35 to an average saving of $128,867.

Table 23: Implied cost saving by shifting a child two NZ Treasury risk factors

Movement of two risk factors	Treasury estimate of average total cost before 21	Treasury estimate of average total cost before 35
Stand takes a 2 to a 0	$79,700	$147,000
Stand takes a 3 to a 1	$76,200	$144,600
Stand takes a 4 to a 2	$48,800	$95,000
Average saving per child	**$68,233**	**$128,867**

Table 23 indicates that for Stand to break even based on it shifting a child by the equivalent of two risk-factors, it must have a lasting positive effect on one in every twelve referred to the Service. That is, $\frac{\$128,867}{10,030} = 12.8$.

[49] If the costs savings to age 21 only are used, Stand must shift one child in three by the equivalent of one Treasury risk factor.

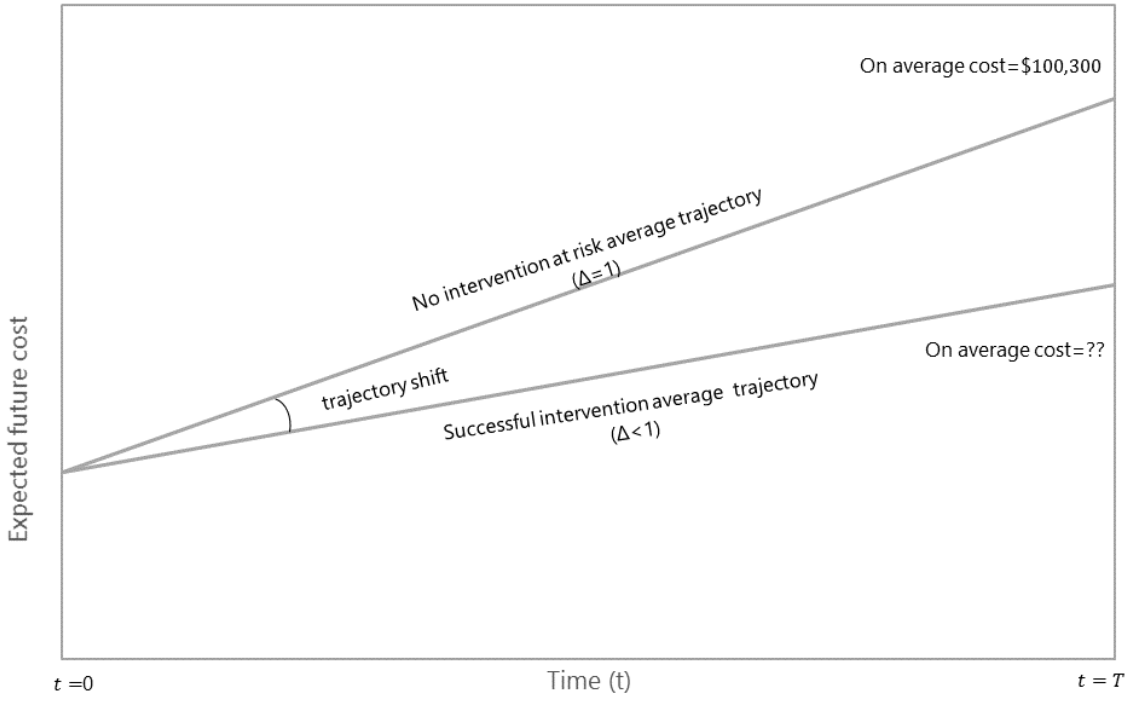
9.2 Summary of fiscal indications and the effect of the Service

This section has presented an introduction to the fiscal implications of the Service. It has demonstrated that if Stand can have an effect that moves a child by effectively one Treasury risk factor, it must have that effect on one in six children for it to be fiscally neutral. That is, if Stand can move one in six children by the equivalent of one Treasury risk factor, the government can expect to recover over time the funding that it provides Stand through reduced spending on welfare, corrections and other related services.

10. An alternative approach to estimating fiscal savings with uncertain outcomes

Stand is effectively focused on children who are currently on a path toward poor life outcomes and altering (where it can) the trajectory of that path to produce better outcomes on average. This report has identified that there is evidence that, on average, the Service has a positive effect on the children it serves, while recognising there is considerable uncertainty about the magnitude of the effect in the longer-term. Figure 21 presents the way the improvement can be thought of in terms of the change in the trajectory of the average child over time.

Figure 21: Graphical representation of trajectory methodology in a point-estimate setting



It is important to note, however, that any trajectory is going to have a large random component as life outcomes are probabilistic functions, and the spread or degree of uncertainty may be large. Focusing solely on the average (or point-estimate) of outcomes doesn't capture the uncertainty and range of possible outcomes.

To better capture the range of possible outcomes, this section present a possible new approach to thinking about the effect and value of Stand's intervention in a distributional setting, rather than a point-estimate setting.

10.1 Defining outcomes as distributional and uncertain

There is a large degree of uncertainty about the future life outcomes of at-risk children. Some children in risky environments will not fall into bad behaviour patterns and create

costs for the government. Other children who are not in at-risk environments will end up in the courts or costing the government in other ways. However, it is known that children who are at-risk or showing risky behaviour are more likely to end up with poor life outcomes. Further, what "poor life outcome" means in terms of the cost to the government will also differ greatly from child to child. The NZ Treasury reports that on average a very risky child can cost the government (up to the age of 35) $320,000, but as reported on the Treasury website, this could be over $1,000,000.[50]

Considering the outcomes as a probabilistic range presents the question of the likely future fiscal outcomes in a different light. If the average cost is $320,000 and some possible maximum is over $1,000,000, what is the distribution of those outcomes and what does that mean about modelling future outcomes and risk?

10.2 The log Normal distribution and Geometric Brownian Motion

In this section, we consider the insights that option pricing and real options analysis can provide.[51] In analysing company stock prices, it is assumed that Central Limit Theorem holds and returns are distributed approximately normally. Also, due to the limited liability of companies, stock prices cannot be negative and returns are compounded through time. If the returns to company shares behave according to a normal distribution, then the prices evolve according to a lognormal distribution which cannot be negative.

Equation 2 below is process of Geometric Brownian Motion (GBM) which is the underpinning assumption for the most common analysis of option and real option pricing. Take X as being the uncertain value of some asset. X therefore, evolves according to,

Equation 2: Geometric Brownian Motion

$$\partial X_t = \mu X_t dt + \sigma X_t W_t$$

where ∂X_t is the change in value at time t to time t+1 over some set time period dt. X_t is the current value at time t. W_t is the Brownian noise term which is an independently and identically distributed (iid) random noise term with a mean of zero and a standard deviation of one (standard normal). Then, σ is the standard deviation of the distribution which determines the size of a movement if there is one. Lastly, μ is the drift term. μ will garner most of our attention because it effectively specifies the trajectory of the asset. That is, if there was no distribution or uncertainty around the future, then over some time-period $t = 0$ to $t = T$, X will become $X_0(1 + \mu)$ as μ is simply the rate of change (like a return).

This approach seems to fit very well the likelihood of outcomes for at-risk children as at each point in life there is some change that is leading toward some terminal value (as indicated by NZ Treasury numbers).

If we know the distribution of costs associated with the life outcomes, then the terminal value can be solved (assuming the GBM process above) by using Ito calculus which gives,

[50] http://www.treasury.govt.nz/statesector/socialinvestment/casestudies
[51] "Investment Under Uncertainty", Avinash Dixit & Robert Pindyck, Princeton University Press, 1994.

Equation 3: GBM solution

$$X_T = X_0 e^{\left(\left(\mu - \frac{\sigma^2}{2}\right)T + \sigma W_t\right)}$$

where X_0 is the value at time 0 ($t = 0$), X_T is the terminal value at the terminal date (in this case 35 years old) and again W_t is the Weiner Brownian noise term.[52,53]

Due to the random nature, it is not possible to have one estimation of this function and conclude anything about it. Therefore, we need to solve for the true distribution. This is done through a Monte Carlo estimation which involves finding a solution with the random term multiple times to get a shape of how the outcomes look, as a whole.

10.3 Outcome estimation approach with uncertainty

If we believe that GBM is a reasonable process for the expected at-risk future costs to follow (and we do) then we need to define what we know and don't know about the distribution and process to which the future fiscal cost evolves before we can start to say anything about it.

10.3.1 What we know

Firstly, we address the terminal value (X_T). Unlike most problems to solve or estimate where we have some current value and an uncertain future value, in this setting we actually have a point estimate of the terminal value that is taken from the Treasury's estimates of the cost of at-risk children to the government. This makes the approach non-standard. However, there is still a solution that will be developed below. The second parameter that we know is the drift term (μ). In the absence of randomness, we know on average X_0 (which represents some ultimately arbitrary present value expected cost at birth) will become X_T which equals the Treasury estimation of at-risk cost. This means that the drift must be equal to one (so $\mu = 1$). Also, we know that the Brownian term is a standard normal "iid" noise term. Lastly, we assume $T = 1$. This is an oversimplification, but we are effectively looking at this as a jump-style process even though it is technically a diffusion. We have no indication of when costs are incurred and the Treasury estimates are as at the ages of 21 and 35 so we believe limiting $T = 1$ is reasonable.

10.3.2 What we don't know

The thing that we don't know is one of the most important in understanding the distribution of any random variable, sigma (σ). Sigma is the standard deviation attached to the point estimate (that we are taking from the Treasury estimates). It does not appear that the Treasury has published the standard deviations attached to its estimates of risk. However, as noted above, on its website the Treasury has stated that for some at-risk classification the average cost to 35 years old is $320,000 and this could be higher than $1,000,000. This gives an indication on the range of figures that are

[52] "The Pricing of Options and Corporate Liabilities", Fischer Black & Myron Scholes, The Journal of Political Economy, 81:3, 637-654, 1973.
[53] "Ito's Calculus and the Derivation of the Black-Scholes Option-Pricing Model", George Chalamandaris & A. (Tassos) G. Malliaris, Handbook of Quantitative Finance, C. F. Lee, Alice C. Lee, eds., Springer, 2009.

experienced in the analysis conducted by the Treasury and again if we believe that changes in the expected cost are approximately normal (implying the total cost are approximately lognormal), then we can choose a sigma that fits this point estimate and range through a grid search.

10.4 Estimating the distribution

First, we need to come up with a process to solve the problem and estimate the distribution. Then it will be possible to see implied changes to the drift and therefore the trajectory on average. This is done in two steps:

1. As we know the point estimate for X_T we need to estimate what X_0 is. This is done through a Monte Carlo estimation that involves several draws of $X_0 = \frac{X_T}{e^{\left(\left(\mu-\frac{\sigma^2}{2}\right)T+\sigma W_t\right)}}$, essentially discounting X_T to X_0 (the number of draws needs to be large enough for the sample of simulated outcomes to converge to the true distribution. For this process, we have adopted 2 million draws which is necessary due to the size of the dollar value involved).

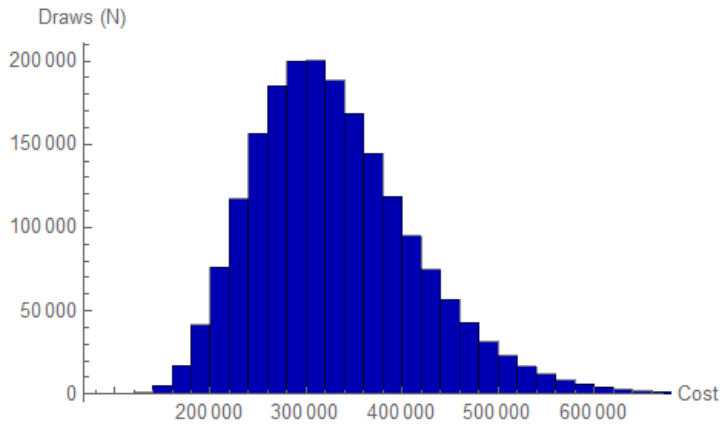
2. We then take the average X_0 from the first Monte Carlo and run a second Monte Carlo to compound X_0 to estimate the distribution forX_T (using $X_T = X_0 e^{\left(\left(\mu-\frac{\sigma^2}{2}\right)T+\sigma W_t\right)}$).

For a more details of the process see Appendix 7.

10.4.1 Estimating sigma

Firstly, we need to estimate a volatility measure as indicated above. To do this we run the procedure outlined above of arbitrary levels of sigma changing by 0.01 each time. We find that the closest fit is when sigma equals 0.18. The two million draws give an average of $320,002 and a maximum observation of $1.08m. This seems to be a reasonable fit to the Treasury estimates of $320,000 and $1m noted above. It gives the distribution shown in Figure 22 below, where the frequency of observations is on the vertical axis and the cost amount is on the horizontal axis.

Figure 22: Implied distribution of Treasury's numbers



10.5 Estimating the distributional effect of a shift in trajectory

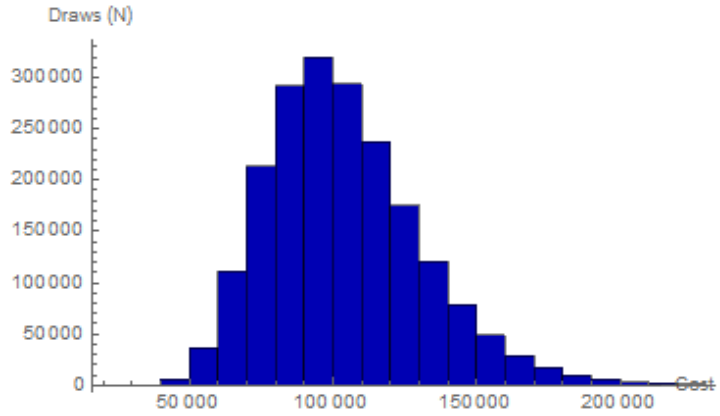
Now that all the parameters have been identified it is possible to estimate the expected distribution and averages for costs of a projected outcome (without intervention) vs. the costs if an effective change in trajectory is achieved.

10.5.1 Moving the needle by one risk factor

To recap, the NZ Treasury has estimated that, on average, a child with one of its defined risk factors will cost the government $100,300 before age 35. We have adopted this number to represent Stand's average client.

Re-running the process defined above, and assuming the fitted sigma is as relevant to the lower risk group as it is to the higher risk group, gives the distribution presented in Figure 23.

Figure 23: Implied distribution for the one-risk-factor group



This simulation (again using two million draws) gives an average cost of $100,311 with a maximum cost of $384,068 and a standard deviation of $26,826. This appears to be an adequate fit for what might be expected.

Now the question of how the trajectory is likely to change must be addressed. Section 8.3 above presented breakdowns of the behavioural profile changes. It showed that, of those who had an improved behavioural profile following Stand's Service, the improvement was approximately a third from the trajectory of that of an at-risk child. In lieu of any better metrics for changed behavioural profile and life outcomes, we adopt this number as a representation of the change in life trajectory. That is, we rerun the estimation process with one important parameter changed. For this estimation, we set the drift term to 0.67 which is 1-0.33 representing a 33% decrease in the poor outcome trajectory ($\mu = 1 - 0.33 = 0.67$).

Figure 24 provides the distribution of the simulated outcomes.

Figure 24: Implied distribution for one-risk-factor group with altered trajectory

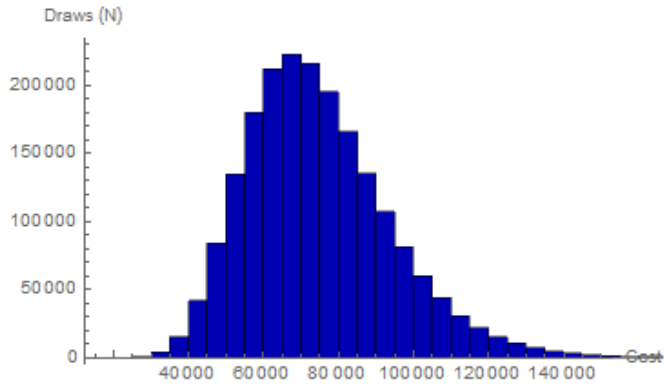


Figure 24 represents the new distribution if a one-third shift in trajectory is achieved by Stand for its clients. This distribution gives an average expected cost for the cohort of $72,089, with a maximum possible cost of $250,065 and a standard deviation of $19,273.

This 33% reduction in the trajectory term is illustrative only and is based on several assumptions. Firstly, we are assuming that there is a linear 1:1 relationship between the decrease in behavioural difficulties and positive life outcomes. This is uncertain and we still don't know what this relationship may actually look like (this is addressed in the next section on further work that could dramatically enhance the economics of social investment returns). Secondly and more importantly, we do not wish to overstate the conclusiveness of the 33% decrease in behaviour. We have taken the number from the average decrease in those who were positively affected by the intervention (slightly over half of all of those admitted). This is a place holder until more work can be conducted into SDQ scores and observed life outcomes. Our aim at this time is to present a possible estimation approach that has the ability to account for the uncertainty of child intervention on future outcomes. The approach considers the distribution and not just the final estimated average to gain a clearer picture of possible outcomes and the value of intervention.

Figure 25 below compares the expected distribution for no trajectory change with the distribution of possible outcomes with an effective change in trajectory.

Figure 25: Implied distributions compared

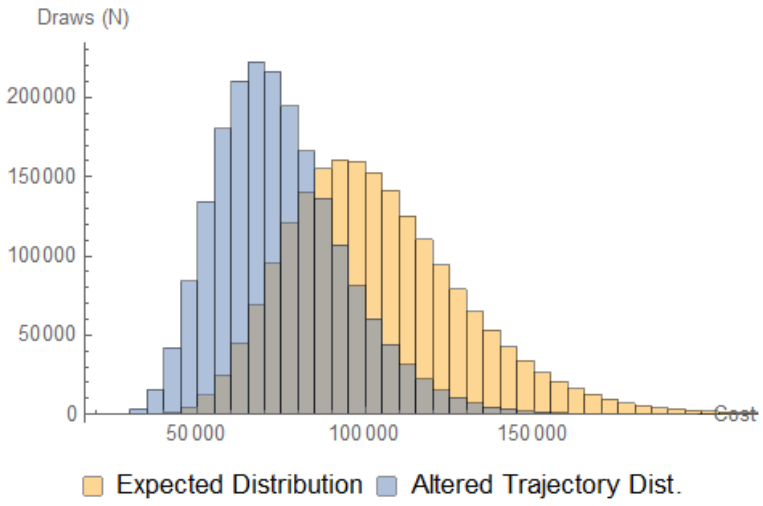


Figure 25 shows the marked difference that can be achieved across the range of possible outcomes if the trajectory toward poor life outcomes can be altered for the better, assuming the average magnitude of the change is one third.

10.6 Summary and possible implications for Stand and its funding for the Service

Due to the assumptions that have had to be made (and in particular the uncertainty around the assumed 33% change in life trajectory) we are limited in what can be said about the fiscal implications of the approach outlined in this section at this stage. The key takeaway from this section of the report is to focus less on change in the average outcome and more on the change in the distribution of outcomes. As depicted in Figure 25 above, improving the expected trajectory of life outcomes has resulted in a tighter distribution, with (on the assumptions used), the standard deviation decreasing in illustrative terms from $26,826 (with an average of $100,300) to a standard deviation of $19,273 (with a mean of $72,089). The analysis highlights that the number of observed lower value outcomes has increased considerably and the number of higher value outcomes has decreased considerably. This means that even though the average has decreased by exactly what might be expected, the probability of outcomes close to the average is higher. In contrast, without the improvement in the expected outcome, the distribution is more dispersed and slightly skewed toward higher value outcomes.

Thus, by improving the average outcome for children-at risk, Stand is also likely to be reducing the range of likely outcomes (as the standard deviation of the distribution declines as the mean declines). Thus, the probability of a child having a very poor outcome declines and the expected outcomes for the worst cases are less severe (and therefore less costly). At the same time the expected outcome for the "average" child improves.

11. Recommendations for further work

The government has initiated some work in recent years to lay the framework for assessing the longer-term fiscal impacts of its social interventions (the "social investment" approach). This section presents some further steps that could be taken to help assess the impact of Stand and other child interventions more generally in the context of the government's social investment framework.

11.1 Data matching in the IDI of Stand's clients and their outcomes

This report has relied heavily on changes in reported behaviour as an indicator of future outcomes from Stand's activities. We have sought to link environmental risk to behaviour and then tracked the changes in behaviour to indicate an effective overall risk of poor life outcomes.

The NZ Treasury has made progress developing a way of assessing the correlation between its defined environmental risk factors and the expected costs to the government in the long term of children exhibiting these factors. While having the Treasury estimates helps to present a general counterfactual for the "at risk" costs, this still leaves the question of what impact an intervention like Stand can be expected to have on the four Treasury risk factors. Many if not most interventions focused on child health and the family environment are unlikely to be able to change any of the four risk factors presented by the Treasury even though the interventions may have an impact on the life outcome of the average child and therefore the long-term costs to the government.

We understand preliminary agreement has been reached to integrate Stand's data into the IDI. We therefore recommend that Treasury, Statistics NZ and Stand seek to develop a clear action plan to achieve this integration in the near future so that Stand's clients can be followed through time to gain an indication of how strong a link there is between Stand's impacts on its clients and fiscal outcomes over time. Further to this, and more generally, the behavioural data collected by the 'B4 School Check' could be used in the same fashion to help develop a more rigorous counterfactual.[54]

11.2 Longer-term assessment of the impact of the Service

Stand currently uses the SDQ to assess children at three stages: pre-intervention, post-intervention (immediate) and six months after intervention. These assessments permit a rigorous assessment of the short-term impact of the Service but leave open the question of the longer-term impact of the Service. We recommend that (subject to resourcing constraints) Stand, in partnership with government consider undertaking also a regular longer-term - say two years after intervention - follow up SDQ of a random sample of children that have been clients of Stand.

[54] While still noting the concerns with the use of the SDQ for the specific 'B4 School Check' programme.

11.3 Other possible approaches to consider for assessment of interventions

There are two measures, based on SDQ assessments developed in the literature, primarily from the UK, that are also worth considering when assessing the feasibility of different interventions or programmes. The first is a more recently developed metric for comparing possible programmes or interventions, called the SDQ Value Added Score. The second is a recent attempt to look at behavioural interventions in the economic setting of Cost-Utility analysis commonly used in health-care economics.

11.3.1 SDQ Value Added Scores

The SDQ Value Added Scores were developed to assess the performance of interventions with what would be expected should no intervention have taken place. One study from 2009[55] shows that using the British Child and Adolescent Mental Health Survey 2004 data, with a control group and a treatment group, the SDQ Value Added Score is effective at measuring the size of the effect of the intervention. We note however, that this study focused on those who were assessed as having a psychiatric disorder and may not be as applicable in this setting. However, the methodology has been peer-reviewed and may be useful for assessing general interventions.

Equation 4 below is the fitted regression from the UK data that has been shown to give a Value Added Score equal to zero with no intervention.

Equation 4: SDQ Value Added Score

$$Value\ Added\ SDQ = 2.3 + 0.8 * T_1Total + 0.2 * T_1Impact - 0.3 * T_1Emotion - T_2Total$$

Where, T_1Total is the Total Difficulties score pre-intervention, $T_1Impact$ is the Impact score at pre-intervention, $T_1Emotion$ is the Emotion score at pre-intervention and T_2Total is the Total Difficulties score post-intervention. Given Equation 4, the Stand for Children Service would receive a raw added value SDQ score of 1.47. This indicates that the Total Difficulties score at post-intervention was 1.47 points lower than what would have been expected without intervention. Taking this score and dividing by 5 (which is the standard deviation in the normative sample taken from the 2009 study) gives an effect size of 0.29. We don't know what would be found from a treatment group in this setting (in the current time-period and within a New Zealand random sample), given the control group is based in the UK and uses old data, so there is not much we can conclude from this other than there has been a positive change for Stand's service compared to a UK based counterfactual of no service.

A more recent 2016 publication uses the Value Added SDQ in assessing the benefits of a family therapy intervention within local authorities' social work programmes in the UK.[56] It finds a case for intervention with young children who are exhibiting significant behavioural issues as the issues typically lead to more complex mental health problems,

[55] "Strengths and Difficulties Questionnaire Added Value Score: evaluating effectiveness in child mental health interventions", Tamsin Ford, Judy Hutchings, Tracey Bywater, Anna Goodman, Robert Goodman, The British Journal of Psychiatry May 2009, 194 (6) 552-558.
[56] "Application of the Strengths and Difficulties Questionnaire Added Value Score in evaluating the effectiveness of Functional Family Therapy within local authority social work services", John Marshall, Russell Hamilton, Nicole Cairns, Child and Adolescent Mental Health, 2016.

state dependence, criminal behaviour and substance abuse. The study finds that the family therapy programme gives an added value SDQ raw score of 2.64 and an effect size score of 0.53 which it deems a moderate change, indicating that the intervention, rather than natural events, is responsible for the positive shift in behaviour. While the finding is larger than seen in Stand's service, the two seem relatively comparable.

11.3.2 Cost-utility estimation

The second recent development shown in the literature for objective economic analysis of child behaviour-based interventions is the mapping of SDQ behaviour scores into utility values. This gives a new way of developing a cost-utility approach to assessing the economic implications of any given child behaviour intervention. A recent 2016 publication[57] provides what appears to be a first consideration of how mapping algorithms from the SDQ to utility-based measures could be utilised in an economic evaluation setting. The study finds that the SDQ and the child health utility measure are related. Using this approach would open the door for new assessment approaches to the economic benefit of interventions. The study notes that future research valuing the changes in SDQ scores would add to the research that it has conducted.

While this work is interesting, it would provide a measure comparing the cost with the utility (in the sense of years-added as is typically seen in drug trials and other parts of health economics) of an intervention programme and not a fiscal valuation of avoided costs to the government.

11.4 Summary of further work recommendations

This section has outlined several possible areas of further research that we believe will add value to the conversation about the true long-term value of childhood intervention programmes. These include integrating Stand's data into the IDI, improving Stand's data collection methods and applying recently developed behavioural metrics that may be useful for comparing interventions.

[57] "Paving the way for the use of the SDQ in economic evaluations of school-based population health interventions: an empirical analysis of the external validity of SDQ mapping algorithms to the CHU9D in an educational setting", Boyer NR, Miller S, Connolly P, McIntosh E, Qual Life Research, 2016, 25(4):913-23.

12. Conclusions

This report has analysed the Stand for Children Service operated by Stand Children's Services, specifically focusing on three key questions:

1. is the service targeting the right children?
2. is the service having a positive impact on the children? and
3. what is that impact worth in terms of fiscal savings over time?

Regarding the first question, we consider Stand successful in targeting children who are at higher risk than normal. We find evidence that Stand's Service is focused in regions that are more likely to have instances of abuse, that the clients it serves have a gender and ethnicity profile similar to that of CYF's Youth Justice and Child Youth and Family admittances and that the Service's clients are likely to come from lower decile schools. We also find that the clients referred to the Service exhibit higher risk behavioural profiles than is expected for the population as a whole and that before the intervention, the behavioural profiles of the Service's clients are comparable to what is seen in the children included in the Growing Up in New Zealand study who were classified as being born into a high-risk environment.

Regarding the second question, we find that, while not all children referred to the service exhibited positive behavioural change, there is on average an observed and meaningful positive shift in the children's behaviour from pre-intervention to post-intervention to a six-month follow-up. Furthermore, this change shifted the Service's clients from being comparable to the high-risk birth cohort in the Growing Up in New Zealand study to being comparable to the medium-risk cohort, on average.

Regarding the third question, the available data does not permit us to make definitive conclusions about the fiscal value of the Service at this stage. Estimated future costs of children in at-risk environments produced by the New Zealand Treasury indicate that the service will "break-even" (ie, generate future fiscal savings for the government that recover current funding on Stand's Service) if the Service has a positive[58] impact on one child in six of all the children referred to it. If the Service can do better than this, it will generate a positive financial return for the government. Once Stand's data is integrated with Statistics NZ's IDI, stronger conclusions on the fiscal impact of the Service will be able to be drawn.

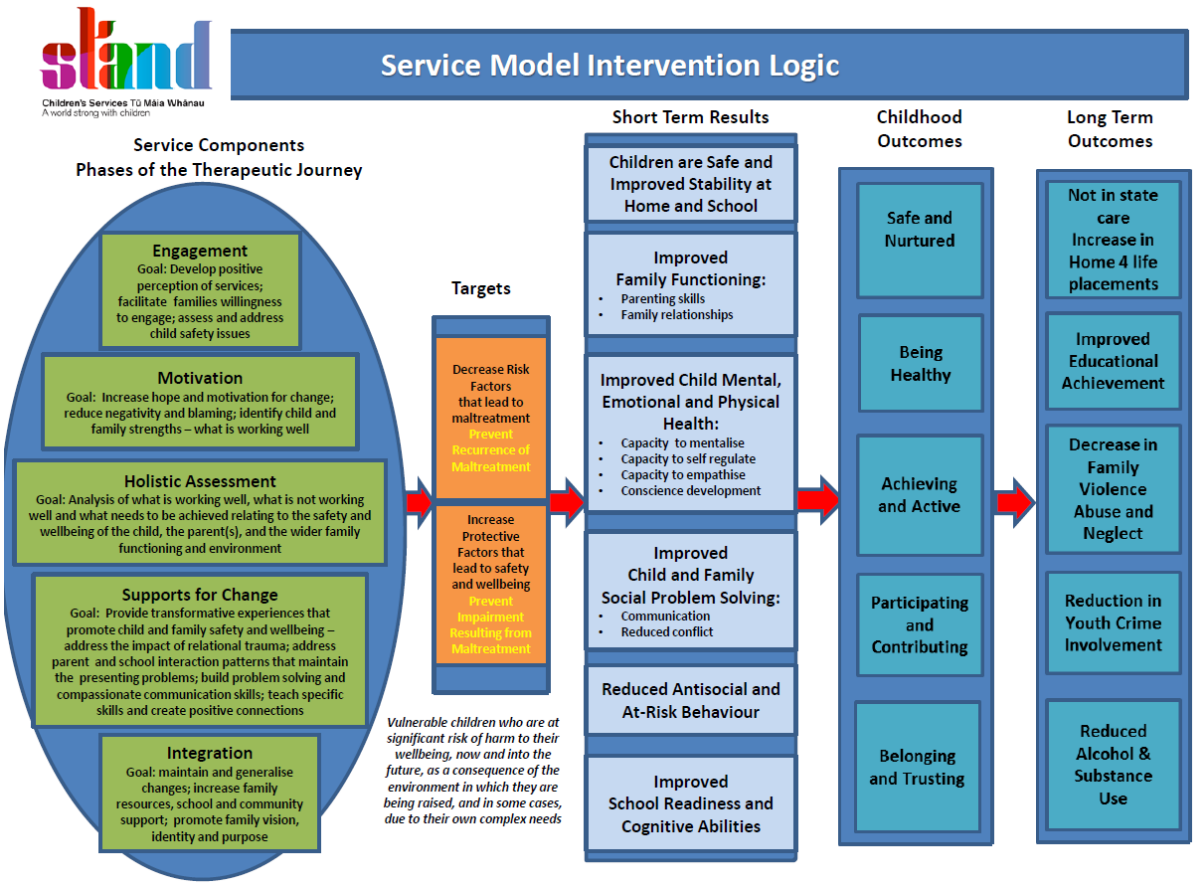
Considering the distribution of future costs to the government of at-risk children highlights that the range of the expected life outcomes of the children Stand serves reduces at the same time as the average expected outcome improves. Thus, following the Service, as well as the expected outcome for an "average" child improving, the probability of a child having a very poor outcome declines and the expected outcomes for the worst cases are less severe (and therefore less costly to the government).

[58] By "positive" we mean decreasing the child's risk profile by effectively one NZ Treasury risk factor on average, as discussed in Section 6 and Section 9 of this report.

An analytical report of this nature carries with it many limitations, primarily based on gaps in the data and the burgeoning nature of studies that have begun to look and think about intervention and how to quantify costs and benefits in the long-term. In particular, it should be noted that the available data on the impacts of Stand is limited to a "before-intervention and after-intervention" methodology (rather than the ideal "with and without" intervention counterfactual test); beyond the six-month post-intervention period, there is little evidence of whether the observed changes in behaviour are lasting; and there is an endogeneity-bias risk in the data we used in this study. However, with the increasing availability of data, such as the micro-data in the IDI, there is considerable opportunity to understand and develop better methods for evaluating Stand's intervention and outcomes in the future.

Appendix 1: The intervention logic for Stand Children's Services

Figure 26: Stand's intervention logic



Appendix 2: Data-cleaning process

The dataset received from Stand contains 23,012 observations of the demographic data and the behavioural constructs. There is a high level of uncertainty about the validity of some of the data points as some SDQ points will be missed because a child is too young or the reviewer has not been able to/forgotten to carry out the assessment. To clean the dataset, first all entries in the behavioural constructs of -1 or NULL were removed. Then, any entries that are outside of the constructs scoring bands were removed as errors. Lastly, any rows that are all zero were taken as a missing observation. The remaining data is taken at face value.

Appendix 3: SDQ assessments of Stand's clients by age and year

Figure 27: Observed frequency of Stand's SDQ responses by age

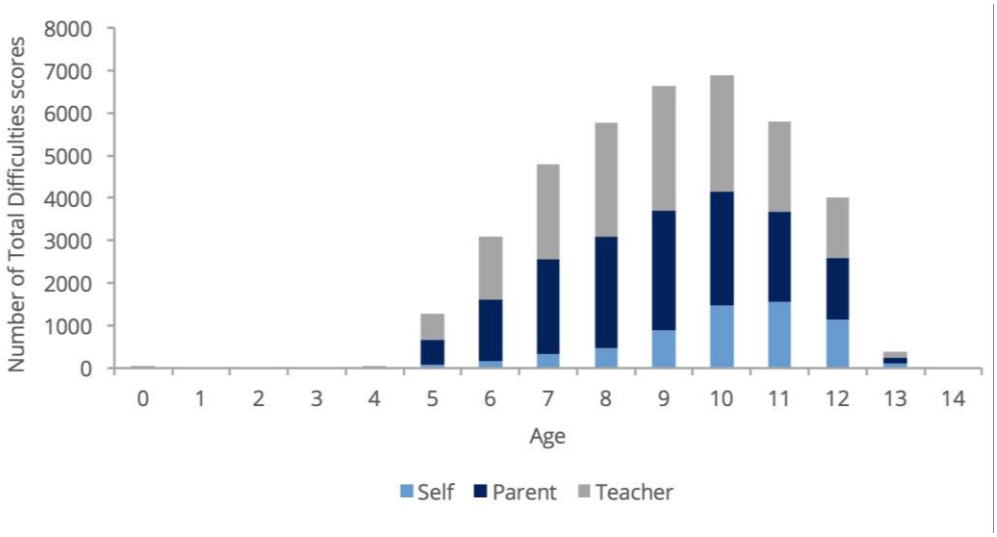
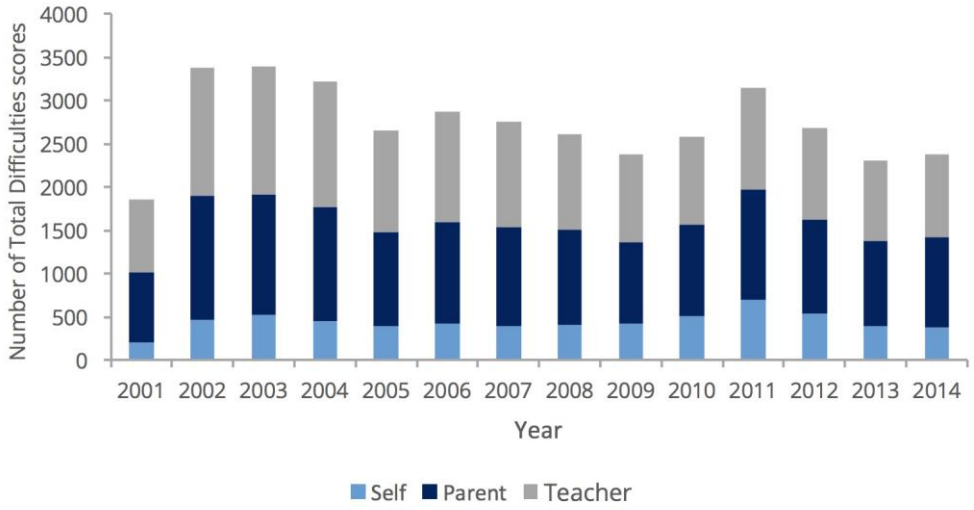


Figure 28: Observed frequency of Stand's SDQ responses by year



Appendix 4: Full breakdown of the normal to abnormal distributions of Stand's clients

Table 24: Detailed pre-intervention SDQ behaviour profile breakdown

	Normal	Borderline	Abnormal	N
Self				
Emotion	72%	12%	16%	6245
Conduct	42%	18%	40%	6242
Hyper	58%	16%	26%	6241
Peer	51%	30%	19%	6244
Pro-Social	77%	11%	12%	6241
Total difficulities	43%	25%	32%	6238
Impact	69%	1%	29%	5761
Parent				
Emotion	49%	14%	37%	16138
Conduct	24%	13%	63%	16140
Hyper	45%	14%	42%	16134
Peer	27%	16%	57%	16141
Pro-Social	69%	16%	16%	16127
Total difficulities	26%	15%	60%	16129
Impact	71%	1%	28%	13698
Teacher				
Emotion	72%	10%	18%	16351
Conduct	42%	11%	46%	16351
Hyper	48%	12%	41%	16350
Peer	47%	14%	39%	16351
Pro-Social	45%	19%	36%	16344
Total difficulities	28%	18%	54%	16344
Impact	67%	1%	32%	15710

Table 25: Detailed post-intervention SDQ behaviour profile breakdown

	Normal	Borderline	Abnormal	N
Self				
Emotion	82%	7%	10%	2378
Conduct	56%	16%	28%	2378
Hyper	67%	15%	19%	2378
Peer	58%	29%	13%	2378
Pro-Social	79%	12%	10%	2370
Total difficulities	57%	21%	21%	2378
Impact	68%	3%	29%	2352
Parent				
Emotion	65%	13%	22%	7872
Conduct	39%	16%	45%	7874
Hyper	61%	12%	28%	7873
Peer	38%	17%	45%	7874
Pro-Social	75%	14%	11%	7861
Total difficulities	47%	15%	38%	7871
Impact	70%	3%	27%	7806
Teacher				
Emotion	78%	9%	14%	7947
Conduct	50%	12%	38%	7948
Hyper	60%	12%	27%	7947
Peer	51%	14%	35%	7946
Pro-Social	50%	18%	32%	7936
Total difficulities	39%	19%	42%	7944
Impact	75%	1%	24%	7854

Table 26: Detailed six-month follow-up SDQ behaviour profile breakdown

	Normal	Borderline	Abnormal	N
Self				
Emotion	74%	8%	17%	362
Conduct	55%	16%	28%	362
Hyper	67%	14%	19%	362
Peer	56%	24%	19%	362
Pro-Social	74%	13%	13%	360
Total difficulities	55%	17%	29%	362
Impact	51%	6%	44%	354
Parent				
Emotion	62%	12%	26%	1,399
Conduct	39%	15%	46%	1,398
Hyper	59%	12%	29%	1,398
Peer	39%	17%	44%	1,399
Pro-Social	73%	14%	12%	1,396
Total difficulities	46%	13%	41%	1,397
Impact	65%	4%	31%	1,377
Teacher				
Emotion	79%	8%	12%	2,584
Conduct	53%	11%	36%	2,584
Hyper	59%	11%	30%	2,584
Peer	56%	13%	31%	2,584
Pro-Social	51%	17%	32%	2,582
Total difficulities	42%	17%	41%	2,584
Impact	73%	1%	26%	2,555

Appendix 5: Supplementary analysis of average scores using a cohort approach

Table 27: Average Total Difficulties scores using different cohort approaches

	Full sample	Cohort all three	Cohort Pre & Post	Cohort Pre & 6m follow-up
		Self		
Average Pre	16.61	15.33	16.37	15.49
N	6238	269	1998	304
Average Post	14.50	14.13	14.48	N/A
N	2378	269	1998	N/A
Average 6m follow-up	14.91	14.86	N/A	14.93
N	362	269	N/A	304
		Parent		
Average Pre	18.00	17.37	17.79	17.32
N	16129	1076	7302	1270
Average Post	14.53	14.21	14.52	N/A
N	7871	1076	7302	N/A
Average 6m follow-up	14.80	14.85	N/A	14.82
N	1397	1076	N/A	1270
		Teacher		
Average Pre	15.94	15.18	15.65	15.25
N	16344	1907	7456	2375
Average Post	13.97	13.54	13.97	N/A
N	7944	1907	7456	N/A
Average 6m follow-up	13.57	13.63	N/A	13.61
N	2584	1907	N/A	2375

Appendix 6: Full breakdown of change directions and magnitudes

Table 28: Impact score direction shifts

Relative to pre-intervention	N Decreased	N No change	N Increased	N	% Decreased	% No Change	% Increased
Post intervention SDQ Total Difficulties							
Self	416	1,112	309	1,837	23%	61%	17%
Parent	1,381	4,342	637	6,360	22%	68%	10%
Teacher	1,228	5,159	840	7,227	17%	71%	12%
Six Month Follow Up SDQ Impact							
Self	122	82	63	267	46%	31%	24%
Parent	290	666	123	1,079	27%	62%	11%
Teacher	401	1,575	311	2,287	18%	69%	14%

Table 29: Impact score magnitude changes: averages and standard deviations

Relative to pre-intervention	Mean decreased	Mean increased	Mean full sample	Standard deviation decreased	Standard deviation increased	Standard deviation full-sample
Post Intervention SDQ Impact						
Self	-72%**	102%**	-41%**	30%	117%	82%
Parent	-73%**	104%**	-55%**	28%	115%	63%
Teacher	-66%**	83%**	-31%**	33%	107%	81%
Six Month Follow Up SDQ Impact						
Self	-74%**	108%**	-51%**	32%	87%	69%
Parent	-70%**	84%**	-49%**	29%	84%	65%
Teacher	-64%**	91%**	-18%**	34%	131%	98%

** statistically different from 0 at $p < 0.01$ level; * statistically different from 0 at $p < 0.05$ level

Table 30: Total Difficulties score direction shifts

Relative to pre-intervention	N Decreased	N No change	N Increased	N	% Decreased	% No Change	% Increased
Post Intervention SDQ Total Difficulities							
Self	1,166	152	680	1,998	58%	8%	34%
Parent	4,822	509	1,971	7,302	66%	7%	27%
Teacher	4,158	511	2,787	7,456	56%	7%	37%
Six Month Follow Up SDQ Total Difficulities							
Self	155	27	122	304	51%	9%	40%
Parent	781	76	413	1,270	61%	6%	33%
Teacher	1,323	148	904	2,375	56%	6%	38%

Table 31: Total Difficulties score magnitude changes: averages and standard deviations

Relative to pre-intervention	Mean decreased	Mean increased	Mean full sample	Standard deviation decreased	Standard deviation increased	Standard deviation full-sample
Post Intervention SDQ Total Difficulities						
Self	-33%**	49%**	-2%	20%	68%	57%
Parent	-36%**	45%**	-11%**	21%	88%	60%
Teacher	-38%**	89%**	11%**	24%	185%	129%
Six Month Follow Up SDQ Total Difficulities						
Self	-35%**	81%**	14%	22%	237%	160%
Parent	-37%**	44%**	-8%**	23%	51%	51%
Teacher	-41%**	98%**	14%**	25%	174%	127%

** statistically different from 0 at $p < 0.01$ level; * statistically different from 0 at $p < 0.05$ level

Table 32: Pro-Social score direction shifts

Relative to pre-intervention	N Decreased	N No change	N Increased	N	% Decreased	% No Change	% Increased
Post Intervention SDQ Pro-Social							
Self	722	445	830	1,997	36%	22%	42%
Parent	2,325	1,675	3,287	7,287	32%	23%	45%
Teacher	2,799	1,313	3,337	7,449	38%	18%	45%
Six Month Follow Up SDQ Pro-Social							
Self	132	65	105	302	44%	22%	35%
Parent	470	262	534	1,266	37%	21%	42%
Teacher	937	369	1,070	2,376	39%	16%	45%

Table 33: Pro-Social score magnitude changes: averages and standard deviations

Relative to pre-intervention	Mean decreased	Mean increased	Mean full sample	Standard deviation decreased	Standard deviation increased	Standard deviation full-sample
Post Intervention SDQ Pro-Social						
Self	-29%**	56%**	12%**	19%	96%	73%
Parent	-28%**	55%**	16%**	18%	79%	65%
Teacher	-39%**	97%**	27%**	23%	126%	105%
Six Month Follow Up SDQ Pro-Social						
Self	-34%**	46%**	1%	20%	82%	61%
Parent	-28%**	51%**	11%**	17%	76%	62%
Teacher	-43%**	98%**	26%**	25%	124%	106%

** statistically different from 0 at $p < 0.01$ level; * statistically different from 0 at $p < 0.05$ level

Table 34: Peer score direction shifts

Relative to pre-intervention	N Decreased	N No change	N Increased	N	% Decreased	% No Change	% Increased
Post Intervention SDQ Peer							
Self	939	382	682	2,003	47%	19%	34%
Parent	3,627	1,504	2,177	7,308	50%	21%	30%
Teacher	3,332	1,438	2,691	7,461	45%	19%	36%
Six Month Follow Up SDQ Peer							
Self	130	53	121	304	43%	17%	40%
Parent	607	270	394	1,271	48%	21%	31%
Teacher	1,136	417	824	2,377	48%	18%	35%

Table 35: Peer score magnitude changes: averages and standard deviations

Relative to pre-intervention	Mean decreased	Mean increased	Mean full sample	Standard deviation decreased	Standard deviation increased	Standard deviation full-sample
Post Intervention SDQ Peer						
Self	-53%**	105%**	6%**	26%	101%	91%
Parent	-51%**	87%**	-2%**	27%	93%	79%
Teacher	-54%**	112%**	10%**	28%	122%	104%
Six Month Follow Up SDQ Peer						
Self	-56%**	115%**	16%**	28%	92%	97%
Parent	-54%**	98%**	1%	28%	107%	90%
Teacher	-57%**	124%**	8%**	28%	128%	110%

** statistically different from 0 at $p < 0.01$ level; * statistically different from 0 at $p < 0.05$ level

Table 36: Hyperactive score direction shifts

Relative to pre-intervention	N Decreased	N No change	N Increased	N	% Decreased	% No Change	% Increased
Post Intervention SDQ Hyperactivity							
Self	954	375	671	2,000	48%	19%	34%
Parent	4,040	1,325	1,941	7,306	55%	18%	27%
Teacher	3,734	1,346	2,382	7,462	50%	18%	32%
Six Month Follow Up SDQ Hyperactivity							
Self	141	51	112	304	46%	17%	37%
Parent	670	231	370	1,271	53%	18%	29%
Teacher	1,148	409	819	2,376	48%	17%	34%

Table 37: Hyperactive score magnitude changes: averages and standard deviations

Relative to pre-intervention	Mean decreased	Mean increased	Mean full sample	Standard deviation decreased	Standard deviation increased	Standard deviation full-sample
Post Intervention SDQ Hyperactivity						
Self	-42%**	87%**	8%**	24%	98%	82%
Parent	-41%**	68%**	-6%**	23%	87%	66%
Teacher	-45%**	91%**	4%**	26%	109%	87%
Six Month Follow Up SDQ Hyperactivity						
Self	-44%**	103%**	17%**	25%	123%	102%
Parent	-42%**	71%**	-2%	24%	84%	69%
Teacher	-47%**	95%**	7%**	26%	114%	93%

** statistically different from 0 at $p < 0.01$ level; * statistically different from 0 at $p < 0.05$ level

Table 38: Conduct score direction shifts

Relative to pre-intervention	N Decreased	N No change	N Increased	N	% Decreased	% No Change	% Increased
Post Intervention SDQ Conduct							
Self	989	399	613	2,001	49%	20%	31%
Parent	4,042	1,385	1,880	7,307	55%	19%	26%
Teacher	3,327	1,720	2,418	7,465	45%	23%	32%
Six Month Follow Up SDQ Conduct							
Self	129	67	108	304	42%	22%	36%
Parent	623	263	384	1,270	49%	21%	30%
Teacher	1,021	531	824	2,376	43%	22%	35%

Table 39: Conduct score magnitude changes: averages and standard deviations

Relative to pre-intervention	Mean decreased	Mean increased	Mean full sample	Standard deviation decreased	Standard deviation increased	Standard deviation full-sample
Post Intervention SDQ Conduct						
Self	-51%**	97%**	1%	25%	103%	86%
Parent	-52%**	82%**	-11%**	26%	88%	73%
Teacher	-62%**	122%**	1%	29%	135%	110%
Six Month Follow Up SDQ Conduct						
Self	-54%**	107%**	10%	26%	128%	103%
Parent	-53%**	90%**	-2%	27%	91%	81%
Teacher	-66%**	131%**	4%	29%	135%	117%

** statistically different from 0 at $p < 0.01$ level; * statistically different from 0 at $p < 0.05$ level

Table 40: Emotion score direction shifts

Relative to pre-intervention	N Decreased	N No change	N Increased		% Decreased	% No Change	% Increased
Post Intervention SDQ Emotion							
Self	1,017	372	614	2,003	51%	19%	31%
Parent	4,009	1,426	1,871	7,306	55%	20%	26%
Teacher	3,330	1,493	2,640	7,463	45%	20%	35%
Six Month Follow Up SDQ Emotion							
Self	148	45	111	304	49%	15%	37%
Parent	674	215	382	1,271	53%	17%	30%
Teacher	1,089	469	819	2,377	46%	20%	34%

Table 41: Emotion score magnitude changes: averages and standard deviations

Relative to pre-intervention	Mean decreased	Mean increased	Mean full sample	Standard deviation decreased	Standard deviation increased	Standard deviation full-sample
Post Intervention Sdq Emotion						
Self	-54%**	107%**	1%	27%	120%	96%
Parent	-62%**	105%**	-14%**	28%	112%	89%
Teacher	-66%**	144%**	5%**	29%	141%	121%
Six Month Follow Up SDQ Emotion						
Self	-56%**	143%**	21%**	28%	166%	136%
Parent	-62%**	98%**	-10%**	29%	102%	90%
Teacher	-70%**	159%**	4%	29%	161%	133%

** statistically different from 0 at $p < 0.01$ level; * statistically different from 0 at $p < 0.05$ level

Appendix 7: More detailed outline of Monte Carlo estimation process

This appendix provides two more technical notes about the Monte Carlo process followed in the report. Firstly, all calculations to estimate the correct distribution had to be made in terms of natural logs due to the nature of the log normal distribution. Secondly, due the fact that the discounted average from step 1 is itself a random variable, allowance had to be made for the volatility in step 2. Solving this, the volatility term for the future estimation becomes the original volatility multiplied by the square root of two ($\sigma_2 = \sigma_1\sqrt{2}$).