



Date: 4 February 2014

The feasibility of using predictive risk modelling to identify new-born children who are high priority for preventive services - companion technical report

Introduction

- 1 This report is a technical companion to the report *The feasibility of using predictive risk modelling to identify new-born children who are high priority for preventive services*. It provides further detail on:
 - data linkage
 - the study population, target outcome variable, and predictor variable definitions
 - the modelling strategy and performance
 - the profile of the population and high risk groups.
- 2 Base SAS version 9.3 is used for data linking and to create variables from linked databases. It was also used to develop and test Multilevel models. The SAS Enterprise Miner version 7.1 is used for all other modelling and for scoring population-level data.

Data linkage

- 3 There is no single unique identifier allocated to an individual against which all their contact with different government services is recorded.¹ In order to link data for the same individual across the different systems, it is necessary to match data using names and dates of birth and other potentially identifying variables.
- 4 The electronic data matching algorithms developed for the project to date involve a number of steps:
 - standardising the data so that it can readily be compared

¹ Even within the Ministry of Social Development, care and protection and youth justice information in the Child Youth and Family CYRAS system is attached to a different unique identifier to that used in the benefit system SWIFTT. And within SWIFTT, the unique identifier applied when a client is an adult is different to that used when they were included in benefit as a child.

- developing dummy records to allow records where nicknames or transliterated names are used for the individual to be matched with those where formal or official names are used and to allow inspection of aliases recorded in the different systems
 - “blocking” the data to be compared to limit the number and computational intensity of the comparisons required
 - developing match criteria to identify and link records within and between systems that can reliably be presumed to relate to the same individual.
- 5 Before linking across systems, we first “de-duplicate” the data by identifying and linking records where an individual has multiple administrative identities within a database so that all of these can be taken into account when describing the individual’s contact with a service and when linking records for them held across systems.
 - 6 Once records are matched, it is possible to compare date of birth and sex information held for an individual across databases. Where there is conflicting information, rules are applied to prioritise the information that appears to be the most reliable. Where demographic information comes from birth registration data, this is given the highest priority as it is considered the most reliable.
 - 7 In the analysis, only ethnicity data recorded on birth registrations is used. This is considered the best quality ethnicity information for the child.
 - 8 At the completion of the data linking and the cleaning of demographic information, each unique individual in the data is given an anonymised research number. The datasets made available for analysis include this anonymised research number, and exclude identifying information such as names, dates of birth and addresses. Careful processes are in place to assure the security of personal data throughout the process of transfer, linkage and analysis.

Match criteria

- 9 The linkage of data formed for the study is “conservative” in that data are only linked if there is a very high level of agreement between identity details in the two systems. The aim is to approximate an administrative approach which would seek to minimise the likelihood of erroneously linking one individual’s information to another individual.
- 10 For children, records are linked within (to de-duplicate) and between data bases where all of the following criteria are met:
 - there is an exact match on the child’s date of birth
 - there is a high level of agreement on the child’s given names (which allows, for example Ann Marie to link with Ann, and Wiremu to link with William)
 - there is a high level of agreement on the child’s surname (which allows, for example Smith-Dickson to link with Smith)
 - there is an exact match on the first given name of at least one of the parents or caregivers.
- 11 Parents and caregivers’ records are linked within and across health, birth, benefit adult, benefit child and care and protection child data where there is:
 - an exact match on the date of birth
 - an exact match on the first given name

- a high level of agreement on the surname (which allows, for example Smith-Dickson to link with Smith, and provides for inspection of both the parent's current name and name at birth recorded in birth registration data, and aliases recorded in the different systems).
- 12 Corrections data for parents and caregivers are linked to records from other data bases where there is:
- an exact match on the date of birth
 - an exact match on the given names
 - an exact match on the surname (which provides for inspection of both current name and name at birth recorded in birth registration data).
- 13 This reflects the very high quality of the Corrections data.
- 14 "Less conservative" linkages are also formed. These allow linkages to be made where the above criteria are not met but the chance of the data relating to the same individual is high. This allows children's records to be linked, for example:
- where they have the same birthdate and similar names (but not sufficiently similar to meet the conservative match criteria)
 - where they have a similar birthdate (or birthdate is missing in care and protection data) but all the other conservative match criteria are met
 - where caregiver information is missing in the care and protection data but all the other conservative match criteria are met.

Linkages formed for modelling

- 15 One of the tasks of the feasibility study is to assess which administrative data are needed in order to make the best predictions of risk, and to assess the sensitivity to the level of conservatism in data linkage.
- 16 To address these questions, five different data linkages are analysed and reported on:
- link 1c draws on benefit and care and protection data only (conservative link)
 - link 2c draws on corrections, benefit and care and protection data (conservative link)
 - link 3c draws on birth, corrections, benefit and care and protection data (conservative link)
 - link 3 draws on birth, corrections, benefit and care and protection data (less conservative link)
 - link 4c draws on health, birth, corrections, benefit and care and protection data.

Assessing the quality of the linked data

- 17 To assess the quality of the linkages formed for the study, a clerical check was undertaken for a stratified random sample of linked records for 527 children born in 2010² and 176 of their parents or caregivers.³

² Records were sampled for children who appeared in each of the administrative systems included in the check within 3 months of birth with the aim of approximating the study population used for this research.

- 18 The checkers examined the quality of the identity linkages across the benefit, care and protection, Corrections and birth registration systems (links 3 and 3c).⁴ Where the data being linked included benefit or care and protection identities, searches were carried out by systems experts using the relevant administrative systems. Where it included other agencies' data, searches for possible missed and false positive matches were conducted within the identity data supplied.
- 19 Table 1 shows that in the vast majority of cases, the linkages formed for the children agreed with those arrived at on clerical inspection, with a higher rate of agreement for the less conservative data linkage (99.6 percent) than for the conservative linkage (95.3 percent). All the errors found involved missed linkages.

Table 1. Proportion of cases in which child linkages agree with clerical check

Group	Conservative linkage (%)	Less conservative linkage (%)
Across all children	95.3 (95% CI 93.1 - 97.0)	99.6 (95% CI 97.5 - 100.0)
Children with findings of maltreatment by age 2	82.4 (95% CI 77.3 - 86.2)	99.0 (95% CI 95.0 - 99.8)
Children with no findings of maltreatment by age 2	95.8 (95% CI 93.5 - 97.5)	99.6 (95% CI 97.4 - 100.0)

Notes: Weighted estimates based on a stratified random sample of children born in 2010 (n = 527)
95% CI = 95 percent confidence interval for the estimates.

- 20 The area that posed the most problems was the linkage of care and protection records for a child with records from other systems, with agreement between the conservative linkage and the clerical check in only 82.4 percent of cases where the child had findings of maltreatment by age two. Virtually all of the unmatched cases could be successfully linked in the less conservative linkage, with agreement between that linkage and the clerical check in 99.6 percent of cases.
- 21 Difficulties in linking care and protection data reflect a general practice of establishing care and protection administrative identities without identity verification. Missed matches frequently occur where information about the child's identity omits or estimates key information (such as date of birth), is subject to some inaccuracy (for example, in the reporting of the child's name), or uses informal rather than formal names for the child, parents or caregivers.
- 22 In contrast, children included in a caregiver's benefit must have their identity verified against a birth certificate, either before or shortly after the granting of the child inclusion. As a result, the child identity information in the benefit system is of relatively good quality.⁵
- 23 When looking at the accuracy of the links for a child and all the adults associated with them, the rate of agreement with links arrived at on clerical inspection was lower at 81.9 percent for the conservative linkage and 93.8 percent for the less conservative linkage (Table 2). The lower accuracy reflects the increased likelihood that any one of the people involved could have had an error in the data linkage, and the higher rate at which errors were found for the adults reflecting their longer, and in some cases more complex, histories.

³ This involved checking linkages for the parents and caregivers of around one in five of the children in each strata.

⁴ At the time the clerical check was performed the project team had not received the Ministry of Health data. As a result, identities from the Health system are not included in the checking reported here.

⁵ However, it is apparent that birth certificates are not always referred to as a small number of children are included in benefit for months or years before their birth is registered. See Figure 5 in the main report.

Table 2. Proportion of cases in which child and all associated adults' linkages agree with clerical check

Group	Conservative linkage (%)	Less conservative linkage (%)
Across all children	81.9 (95% CI 67.7 - 91.7)	93.8 (95% CI 80.4 – 98.8)

Notes: Weighted estimates based on a stratified random sample of children born in 2010 (n = 96).

Includes checks of records for 176 parents or caregivers.

95% CI = 95 percent confidence interval for the estimates.

- 24 For the adults, all of the errors found in the conservative linkage were missed matches. In the less conservative linkage, the error rate for adults was lower, but errors found included false positive linkages where one individual's information was erroneously linked to another individual. (In the less conservative linkage, false positive linkages were found for four of the 176 adult linkages checked, and missed matches were found for two of the 176 adult linkages checked).
- 25 Depending on the number of other children in the family, the analysis data for one child could potentially draw on as many as 20 to 30 individual-to-individual linkages between systems. The estimates presented here (which only consider accuracy of linkages for the reference child and the related adults) should therefore be viewed as conservative. In practice, the proportion of cases in which the information informing risk scoring might, under a conservative data linkage, be incomplete as a result of missed linkages would be higher than indicated in Table 2.
- 26 While for the children no false positive matches were found in the clerical review, these were uncovered in other examination of the data. In some of these cases, information for twins was erroneously linked due to the similar patterning of the two children's names (for example, twins with the same given name, different middle names, and same surname were erroneously linked and treated as one child under both conservative and less conservative linkages).

Study population

- 27 The population considered varies depending on the linkage of data used to develop the model.
- 28 In the base models, we draw on linked births, corrections, benefit and care and protection data (link 3c) and consider all children in a birth cohort who are either:
 - identified through birth data by three months of age (ie. a birth registered within 91 days of birth), where the child was not still-born, or
 - identified through benefit data by three months of age (ie. included in a main welfare benefit within 91 days of birth).
- 29 With inclusion of health data, we develop and test models that draw on linked health, births, corrections, benefit and care and protection data (link 4c) and include in the base model additional and augmented predictor variables. These models consider the same study population as the base models.
- 30 Two models are developed and tested on more restricted data linkages (1c and 2c). These consider only children identified through benefit data by three months of age.
- 31 The three month window could be altered. The aim is to demonstrate a hypothetical predictive risk modelling (PRM) tool that would draw on administrative data available shortly after birth.

- 32 Generally, we develop models for the 2010 birth cohort and test the model performance on the 2007 birth cohort (which allows a longer window for follow up in order to assess the predictive accuracy of the models).
- 33 A descriptive profile of these cohorts for link 3c is provided in Table 3 below. More detail on variable definitions and derivation is provided in Appendix 1. Note that because the profile is based on conservatively linked data, this will understate the proportions with care and protection history and with findings of maltreatment.
- 34 For more recent cohorts, the care and protection history of caregivers in their childhood is more complete and this mainly explains the differences in the profile of birth cohorts 2007 and 2010 for this variable.
- 35 The proportion of cases where there are other children with a care and protection history is greater for the 2010 cohort. This is consistent with a general increase in contact with care and protection services associated with increased public awareness and changes in Police notifications for family violence which saw increased numbers of children with findings of emotional abuse (cases where there is a Police family violence notification or contact record but no other notification and no findings are treated as no history in the derivation of the variable).
- 36 Comparing the 2007 and 2010 cohorts, there was an increase in the proportion of children with caregivers with a history of receiving benefit for a substance abuse or mental health disorder in the last five years. This is consistent with a long-term trend of increasing rates of receipt of incapacity related benefits (Ministry of Social Development, 2012).

Table 3 (a). Profile and cumulative incidence of substantiated findings of maltreatment by age 2, children in the study population born in 2007 and 2010 (link 3c)

		Cohort 2007 (N=62273)		Cohort 2010 (N=63176)	
		% in population	Incidence of finding %	% in population	Incidence of finding %
Gender of child	Male	51.5	2.4	51.3	2.8
	Female	48.5	2.2	48.7	2.7
Low birth weight or pre-term	Yes	8.2	3.0	8.4	3.3
	No or unknown	91.8	2.2	91.6	2.7
Parenting demands	High parenting demands	20.1	3.8	20.8	4.6
	No other children	59.4	2.0	49.9	2.4
	Other children but not high parenting demands	20.5	1.8	29.3	2.1
Other children with care and protection history	Yes	4.8	16.9	7.0	16.9
	No	95.3	1.6	93.0	1.7
Family violence	Events in one month	0.6	22.8	1.5	18.9
	Events in more than one month	0.2	36.4	0.8	26.9
	No events (no Police FV notifications or contact records)	99.3	2.1	97.8	2.4
Caregiver's age	Under 20	7.8	6.6	7.2	8.3
	20 to 25	17.5	4.1	18.9	5.1
	25 to 30	24.0	2.0	24.8	2.2
	30 to 35 (includes missing)	28.6	1.1	27.5	1.4
	35 to 40	18.1	1.1	17.3	1.2
Over 40		4.0	2.1	4.5	2.2
Benefit caregiver is not a birth registration parent	Yes	0.8	6.7	0.8	8.6
	No birth registration	6.3	9.6	7.3	10.1
	No	92.9	1.8	91.9	2.1
Single parent	Single parent	19.4	7.0	22.1	8.2
	Single parent and no father listed on birth registration	5.2	6.7	4.5	6.5
	Not single parent or partnership status unknown	75.4	0.8	73.4	0.9
Time on benefit in the last 5 years	More than 80%	11.7	10.0	10.7	11.3
	20<-80%	17.0	4.2	17.1	6.6
	Up to 20%	14.1	1.8	13.5	2.1
	No time	57.1	0.3	58.7	0.3
Caregiver with care and protection history	Yes	9.8	10.1	13.1	11.2
	No	90.2	1.5	86.9	1.5
Benefit address changes in the last year	No address changes	23.0	1.8	20.0	2.0
	1-2 address changes	9.8	3.8	9.7	4.9
	3 plus address changes	1.8	12.4	1.9	12.9
	Missing (no benefit in last year)	65.4	2.0	68.5	2.4
Mental health in the last 5 years	Substance abuse issues	1.1	12.5	1.3	14.5
	Persistent substance abuse issues (3+ years in last 5)	0.4	14.3	0.5	19.6
	Mental health issues other than substance abuse	3.6	6.6	4.1	7.3
	Persistent mental health issues other than substance abuse (3+ years in last 5)	1.3	12.2	1.7	11.7
	No known mental health or substance abuse issues	93.6	1.8	92.3	2.1
Behavioural or relationship difficulties as a child	Yes	3.5	12.4	4.3	14.9
	No	96.5	1.9	95.7	2.2
Caregivers' Corrections history in the last 5 years	Non-custodial sentence	4.1	8.2	4.5	9.4
	Custodial sentence for non-violent crimes	1.5	6.9	1.4	10.0
	Custodial sentence for violent crimes	1.7	9.3	2.0	11.1
	No history	92.8	1.9	92.2	2.2

Table 3 (b). Profile and cumulative incidence of substantiated findings of maltreatment by age 2, Māori children in the study population born in 2007 and 2010 (link 3c)

		Cohort 2007 (N=17,288)		Cohort 2010 (N=17,346)	
		% in population	Incidence of finding %	% in population	Incidence of finding %
Gender of child	Male	52.1	5.1	51.6	6.3
	Female	47.9	4.7	48.4	6.0
Low birth weight or pre-term	Yes	8.4	6.1	9.1	7.6
	No or unknown	91.6	4.8	90.9	6.0
Parenting demands	High parenting demands	24.2	7.4	25.0	8.7
	No other children	57.4	4.1	46.6	5.7
	Other children but not high parenting demands	18.4	4.2	28.4	4.8
Other children with care and protection history	Yes	9.7	19.5	13.7	20.2
	No	90.3	3.3	86.3	3.9
Family violence	Events in one month	1.2	26.5	2.9	22.7
	Events in more than one month	0.4	39.1	1.6	29.8
	No events (no Police FV notifications or contact records)	98.4	4.5	95.5	5.3
Caregiver's age	Under 20	16.3	7.0	14.1	9.6
	20 to 25	27.0	5.8	29.1	7.4
	25 to 30	23.4	4.3	23.5	5.5
	30 to 35 (includes missing)	19.4	3.3	18.8	4.1
	35 to 40	11.0	6.2	10.9	3.4
Benefit caregiver is not a birth registration parent	Over 40	2.8	3.4	3.6	6.1
	Yes	1.5	6.3	1.4	13.3
	No birth registration	9.9	14.3	9.6	15.6
Single parent	No	88.6	3.8	89.1	5.0
	Single parent	36.8	8.9	39.2	10.7
	Single parent and no father listed on birth registration	9.8	7.4	8.6	7.9
Time on benefit in the last 5 years	Not single parent or partnership status unknown	53.4	1.7	52.2	2.5
	More than 80%	24.7	11.3	21.6	13.2
	20<-80%	29.1	5.5	29.9	8.5
	Up to 20%	17.0	2.3	16.9	3.3
Caregiver with care and protection history	No time	29.3	0.5	31.6	0.7
	Yes	20.4	10.9	26.5	12.9
Benefit address changes in the last year	No	79.6	3.4	73.5	3.7
	No address changes	20.0	4.2	17.9	4.8
	1-2 address changes	12.3	6.4	11.8	8.6
	3 plus address changes	3.3	15.0	3.1	16.4
Mental health in the last 5 years	Missing (no benefit in last year)	64.4	4.3	67.3	5.6
	Substance abuse issues	2.2	14.8	2.6	15.6
	Persistent substance abuse issues (3+ years in last 5)	0.6	18.9	1.0	21.0
	Mental health issues other than substance abuse	5.6	8.2	6.1	10.2
	Persistent mental health issues other than substance abuse (3+ years in last 5)	1.7	17.3	2.4	16.9
Behavioural or relationship difficulties as a child	No known mental health or substance abuse issues	90.0	4.1	87.9	5.1
	Yes	7.3	13.7	8.5	16.5
Caregivers' Corrections history in the last 5 years	No	92.7	4.2	91.5	5.2
	Non-custodial sentence	8.1	10.3	8.8	11.5
	Custodial sentence for non-violent crimes	2.8	8.6	2.7	13.5
	Custodial sentence for violent crimes	3.4	10.6	3.9	12.9
	No history	85.7	4.1	84.6	5.1

Target outcome variables

- 37 For base models and the majority of models tested in this study, the target outcome variable considered is any substantiated findings of maltreatment by age two. This is as recorded in care and protection administrative data and includes any substantiated findings of neglect, emotional abuse, physical abuse or sexual abuse.
- 38 The findings date (the date on which findings were determined) is used to identify findings by age two. Generally, findings are recorded within two to three months of notification. In some cases, a notification will have been received by a given age but maltreatment findings will be yet to be determined and recorded.
- 39 In addition, as a part of sensitivity testing, we examine whether and how well the predictor variables that have been identified can predict other target outcome variables. These include the following:
- notifications (including Police Family Violence (FV) notifications and contact records)
 - investigations or Child and Family Assessments
 - substantiated findings of physical or sexual abuse or neglect (excluding emotional abuse findings from the target)
 - substantiated findings of physical abuse or hospitalisation for maltreatment or marker injuries.⁶
- 40 The predictive performance of various models targeting different outcomes will be discussed in later sections of this Companion paper. Table 4 below presents the outcomes measured for cohort 2007 and 2010 by age 2.

Table 4: Outcomes for study population born in 2007 and 2010 (link 3c and 4c)

Outcomes by age 2	Cohort 2007			Cohort 2010		
	All	Incidence	Incidence rate	All	Incidence	Incidence rate
Notifications (including Police FV notifications and contact records)		4,605	7.40%		5,055	8.00%
Investigations or Child and Family Assessments	62,273	2,844	4.60%	63,176	3,463	5.50%
Substantiated findings of maltreatment		1,434	2.30%		1,750	2.80%
Substantiated findings of physical or sexual abuse or neglect (excluding emotional abuse findings)		619	1.00%		683	1.10%
Substantiated findings of physical abuse or maltreatment or marker injury hospitalisation	62,295	191	0.30%	63,200	232	0.40%

Notes: Cohort size varies slightly depending on the link used. Substantiated findings of physical abuse or maltreatment or marker injury hospitalisation is based on link 4c and all other outcomes are based on link 3c. Because it is based on conservatively linked data, this table understates the proportions with different measures of care and protection contact.

- 41 Based on conservatively linked data, about 8.0% of the cohort born in 2010 had a notification, 5.5% had an investigation or Child Family Assessment (CFA) and 2.8% had

⁶ Includes maltreatment injury hospitalisations by age 2 and hospitalisation for intracranial injury or long-bone fracture by age 1. Excludes short-stay emergency department hospital admissions.

substantiated findings of maltreatment by age 2. These rates are slightly lower for the cohort born in 2007.

Predictor variables

- 42 Table 5 provides a list of predictor variables included in models using different linkages. Appendix 1 provides corresponding long descriptions, details of the derivation of each, and limitations.
- 43 The base (population-wide) models based on link 3c and 3 consider 15 predictor variables of which only one (CYF service centre) is a local level variable. We have constructed alternative local level variables and assessed whether they increase performance of the model. This analysis is presented in a separate section on the handling of local level variables.
- 44 In the model where health data is included (link 4c), two predictor variables are added (these indicate the mother's smoking status and whether there is a previous child for whom no immunisation stage is complete by age one and there is no indication that the parent or caregiver has declined immunisation) and five variables are augmented with the addition of health data (caregiver's mental health in last 5 years, family violence, parenting demands, caregiver's address changes in last year and low birth weight or pre-term birth).

Table 5: Predictor variables included in the models

Variable short description	Linkage of data				
	1c	2c	3c	3	4c
Gender of child	✓	✓	✓	✓	✓
Low birth weight or pre-term			✓	✓	✓
Parenting demands	✓	✓	✓	✓	✓
Other children with care and protection history	✓	✓	✓	✓	✓
Family violence	✓	✓	✓	✓	✓
Caregiver's age	✓	✓	✓	✓	✓
Benefit caregiver is not a birth registration parent			✓	✓	✓
Single parent	✓	✓	✓	✓	✓
Caregiver's highest educational qualification	✓	✓			
Time on benefit in the last 5 years	✓	✓	✓	✓	✓
Caregivers with care and protection history	✓	✓	✓	✓	✓
Caregiver's address changes in the last year	✓	✓	✓	✓	✓
Mental health in the last 5 years	✓	✓	✓	✓	✓
Other children immunisation history					✓
Mother is a smoker					✓
Behavioural or relationship difficulties as a child	✓	✓	✓	✓	✓
Corrections history in the last 5 years		✓	✓	✓	✓
Benefit – prison transitions	✓				
CYF service centre	✓	✓	✓	✓	✓
Total number of variables in the model	14	14	15	15	17

Notes:

1c- linked benefit and care and protection data (conservative link)

2c- linked benefit, Corrections and care and protection data (conservative link)

3c- linked birth, benefit, Corrections and care and protection data (conservative link)

3 - linked birth, benefit, Corrections and care and protection data (less conservative link)

4c- lined health, birth, benefit, Corrections and care and protection data (conservative link)

- 45 Highest educational qualification is only able to be captured for those with previous benefit receipt and was found to have too many cases of missing information, even

within the population known to the benefit system, to be of use in population-wide models.

- 46 Although the source administrative data contains continuous variables, after descriptive analysis of the data and a review of other studies, it was decided to band these and create either binary or categorical variables. It was also considered that a model with binary and categorical variables only would be more stable, and easier to explain to stakeholders and to implement.
- 47 Categorical and binary variables derived from continuous variables include caregiver's age, low birth weight or preterm, number of benefit address changes in the last year, time on benefit in the last 5 years, number of other children identified to caregivers that has been used to create a variable of parenting demands.
- 48 In order to understand the relationship between predictor variables and the outcome variable, we run simple logistic regressions⁷ between each predictor variable and outcome measured (substantiated findings of maltreatment by age two for both link 3c and 4c). This provides information about the strength of association and the variance explained for each predictor variable independent from other predictors.
- 49 The model chi-square test or likelihood ratio test provides a significance test for logistic model. The chi-square and its significance value (p) are presented in the table 6 below. Most of the variables except low birth weight and child's gender have statistically significant association ($p < 0.01$) with dependent variable.

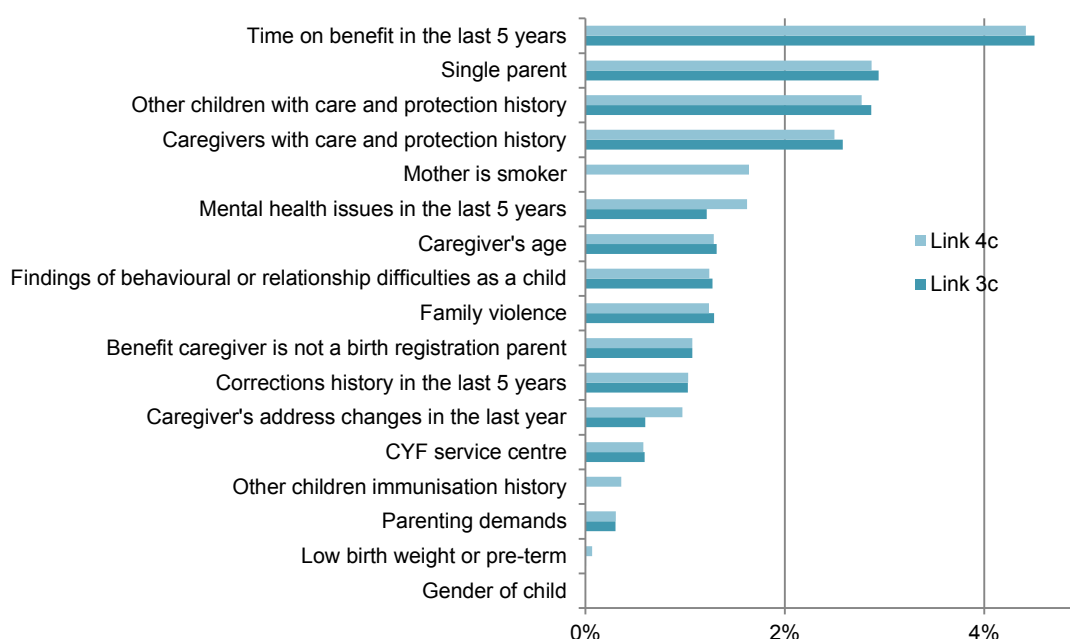
Table 6. Results of likelihood ratio test between each predictor and dependent variable with and without health data (link 3c and 4c).

	Link 3c			Link 4c		
	DF	Chi-Square	Pr > Chi-Square	DF	Chi-Square	Pr > Chi-Square
Gender of child	1	0.3	0.6112	1	0.4	0.5236
Low birth weight or pre-term	1	4.7	0.0304	1	43.7	<.0001
Parenting demands	2	192	<.0001	2	205.7	<.0001
Other children with care and protection history	1	1838.5	<.0001	1	1776.8	<.0001
Family violence	2	822.2	<.0001	2	900.4	<.0001
Age of caregiver	5	839	<.0001	5	820.3	<.0001
Single parent	2	1886.2	<.0001	2	1840.8	<.0001
Time on benefit in the last 5 years	3	2913.4	<.0001	3	2857.1	<.0001
Caregiver with care and protection history as a child	1	1652	<.0001	1	1599.9	<.0001
Caregiver's address changes in the last year	3	381.3	<.0001	3	618.6	<.0001
Caregivers' known mental health issues in the last 5 years	4	775.2	<.0001	4	1033.1	<.0001
Caregivers with findings of behavioural or relationship difficulties as a child	1	810.4	<.0001	1	790.9	<.0001
Caregivers' Corrections history in the last 5 years	3	653.2	<.0001	3	656.1	<.0001
At least one benefit caregiver is not a birth registration parent	2	681.7	<.0001	2	681.3	<.0001
CYF service centre	42	376.8	<.0001	42	370.2	<.0001
Previous child not fully immunised and not declined				1	228.9	<.0001
Mother is a smoker		na		1	1046.7	<.0001

⁷ Logistic regression with one explanatory variable.

50 Pseudo R-square is an Aldrich and Nelson's coefficient which serves as an analog to the squared contingency coefficient with an interpretation like R-square and it can be used to test the strength of association between independent variables and binary dependent variable. The Pseudo R square is presented in Figure 1. On their own, mother or primary benefit caregiver's time on benefit in the last 5 years, relationship status of caregiver, caregiver's care and protection history as a child and care and protection history of other children have greatest association with dependent variable. Child gender and low birth weight or pre-term variables, on their own, have weakest association with dependent variable.

Figure 1. The variance explained (R square) of each predictor variable in isolation with and without health data (link 3c and 4c)



Notes:

3c- linked birth, benefit, Corrections and care and protection data (conservative link)

4c- lined health, birth, benefit, Corrections and care and protection data (conservative link)

51 In the final models, the parameter estimates of each predictor variable indicate the predictive ability of each variable once the effects of other predictors are controlled.

Modelling strategy

Datasets

52 Depending on the research questions raised, two types of datasets are created: (i) sample datasets - the datasets used to develop and test various versions of the models

and (ii) datasets for scoring- the datasets containing population level data on cohorts 2007 and 2010 that are used to assess the accuracy and proportionality of the scoring of the models. Table 7 below lists all sample and scored datasets used in this study. The method used to construct sample datasets is set out below.

Sampling method

- 53 At the population level, we are seeking to model the occurrence of events that are rare. A common approach for dealing with prediction of rare events and associated class imbalance in the source data is to change class distribution and create more balanced data for modelling. Studies have shown that a balanced dataset provides improved classification performance as compared with an imbalanced dataset and under sampling and over-sampling are very effective methods for dealing with the class imbalance problem (Japkowicz, 2000; Laurikkala, 2001 and Estabrooks et al., 2004; Weiss, 2004; Choi, 2010). In order to improve the model accuracy, sampling was carried out so that models draw on most informative data. In the base models that draw on link 3c we under-sample the majority class (no maltreatment findings cases), while capturing all of the maltreatment findings cases.
- 54 The following rules are applied for sampling. The sample includes all children in the cohort who have findings of maltreatment by age two and the ratio between event and no event groups in the sample is kept at 1/5. This means that 20 percent of the sample comprises records for children with substantiated findings of maltreatment by age two and 80 percent comprises records for children with no substantiated findings of maltreatment by that age. SAS Enterprise Miner can adjust the predicted probabilities taking into account the distribution of events for the entire population.
- 55 There were a few exceptions to the sampling rule above.
- In models that draw on benefit, care and protection and corrections data (link 2c and 1c) and the separate models for children of beneficiaries (link 3c and 4c), the target outcome is not rare and it is not necessary to sample to adjust for imbalance - the sample size for modelling is the same as the number in the cohort study population. Similarly, in the model to predict notifications, no sampling was carried out.
 - In the model to predict the substantiated physical abuse or maltreatment or marker injury hospitalisations and the model for children of non-beneficiaries (link 3c and link 4c), we pooled together data for three cohorts (2008, 2009 and 2010) before applying the same sampling rule because of very small numbers with the outcomes targeted.
- 56 Table 6 shows the sample size of each dataset used for modelling and their share in the population level data from which the sample is drawn.
- 57 Throughout the analysis we systematically assess the models' predictive performance by applying them to entire 2007 birth cohort (without sampling).

Data partitioning

- 58 All sample datasets used for modelling are partitioned using the 70/30 rule, where the training dataset contains 70 percent of data and the model is validated on the remaining 30 percent of records. This is a standard approach in data mining (Witten et al., 2011; Williams, 2011). A stratified partitioning method is used to ensure both the training and validation datasets have the same proportion of events. For example, the sample dataset on which the 2010 base model is developed (N=8,750) is split into training

(N=6,123) and validation datasets (N=2,627), and in both the training and validation datasets 20 percent of records are for children with substantiated findings of maltreatment by age two.

Table 7. Datasets used for modelling and scoring

Linkage	Model description	Population	Modelling dataset				Scored dataset
			Total sample	% of population	Training 70%	Validation 30%	
3c	Base model for cohort 2010	63,176	8,750	14%	6,123	2,627	62,273
Cohort sensitivity analysis							
3c	Model for cohort 2009	62,639	8,260	13%	5,780	2,480	62,273
	Model for cohort 2008	62,900	7,865	13%	5,504	2,361	62,273
	Model for cohort 2007	62,273	7,170	12%	5,018	2,152	62,273
PRM sensitivity to various administrative data and linkages							
1c	Model based on benefit and care protection data only	16,946	16,944	100%	11,860	5,084	62,273
2c	Model based on benefit, care and protection data and Corrections data	16,932	16,931	100%	11,851	5,080	62,273
4c	Model based on benefit, care and protection, Corrections and Health data	63,200	8,600	14%	6,018	2,582	62,273
3	Model for cohort 2010 using less conservative link 3	62,017	11,075	18%	7,752	3,323	62,273
Separate models to correct for over-representation							
3c	Model where other children with history of contact with CYF	4,448	4,448	100%	3,112	1,336	2,957
	Model where no other children with history of contact with CYF	58,728	5,000	9%	3,498	1,502	59,316
	Model for children of benefit caregivers based on link 3c	16,929	16,929	100%	11,849	5,080	14,441
	Model for children of non-benefit caregivers based on link 3c	140,036	4,340	3%	3,037	1,303	47,832
	Model for Maori ethnic group	17,346	5,340	31%	3,736	1,604	17,288
	Model for "Non-Maori" and ethnic group not recorded	45,830	3,410	7%	2,385	1,025	44,985
4c	Model for children of benefit caregivers based on link 4c	16,932	16,932	100%	11,852	5,080	14,447
	Model for children of non- benefit caregivers based on link 4c	140,088	4,255	3%	2,978	1,277	47,848

Table 7. Datasets used for modelling and scoring- continued

Linkage	Model description	Population	Modelling dataset				Scored dataset
			Total sample	% of population	Training 70%	Validation 30%	
Modelling for various outcomes							
3c	Model to predict notifications (including Police FV notifications and contact records)	63,176	63,176	100%	44,222	18,954	62,273
	Model to predict investigations or Child and Family Assessments	63,176	17,326	27%	12,127	5,199	62,273
4c	Model to predict substantiated physical or sexual abuse or neglect (excluding emotional abuse findings)	63,176	3,415	5%	2,389	1,026	62,273
	Model to predict substantiated physical or maltreatment or marker injury hospitalisation	188,783	3,565	2%	2,495	1,070	62,295
Sensitivity to local level variables							
3c	Model with no local variables	63,176	8,750	14%	6,123	2,627	62,273
	Model with aggregated local level variables	63,176	8,750	14%	6,123	2,627	62,273
	Model with aggregated local level variables based on multiple counts of notifications and investigations	63,176	8,750	14%	6,123	2,627	62,275
Sensitivity to sampling							
3c	Base model with no sampling carried out	63,176	63,176	100%	44,222	18,954	no scoring was carried out
	Base model with sampling carried out on pooled cohort data	63,176	17,010	27%	11,906	5,104	
	Base model with random sampling carried out	63,176	8,760	14%	6,130	2,630	
	Base model with data partitioning 50/50 to training and validation datasets	63,176	8,750	14%	4,373	4,377	
	Base model with data partitioning 60/40 to training and validation datasets	63,176	8,750	14%	5,248	3,502	
	Base model with data partitioning 80/20 to training and validation datasets	63,176	8,750	14%	6,998	1,752	
	Base model with simple random partitioning	63,176	8,750	14%	6,125	2,625	

Notes:

- 1c- linked benefit and care and protection data (conservative link)
- 2c- linked benefit, Corrections and care and protection data (conservative link)
- 3c- linked birth, benefit, Corrections and care and protection data (conservative link)
- 3 - linked birth, benefit, Corrections and care and protection data (less conservative link)
- 4c- lined health, birth, benefit, Corrections and care and protection data (conservative link)

Algorithms tested

59 A range of modelling algorithms were tested and compared for the 2010 birth cohort sample. In total, 12 algorithms were tested. These included:

- Gradient boosting
- DMINE regression
- Neural network with 3 hidden units based on 15 variables
- Neural network with 3 hidden units based on some selected variables
- Neural network with 4 hidden units based on some selected variables
- Partial Least Squares
- Full multiple regression
- Stepwise regression (entry criteria $p < 1.0$ and stay criteria $p < 0.2$)
- Backward elimination regression
- Decision tree with maximum of 2 branches and depth of 6
- Decision tree with maximum of 6 branches and depth of 6
- Multilevel model or Hierarchical Linear Model (HLM).

60 In this section we are assessing the overall performance of above algorithms by comparing the area under the Receiver Operating Characteristics curve (AUR)⁸, misclassification rate⁹ and cumulative lift¹⁰ at 3 percent threshold across validation and scored datasets. Other diagnostics such as positive predictive value (PPV)¹¹ and sensitivity¹² at 3 percent threshold for scored population are provided in Appendix 2 of this report.

⁸ Area under the Receiver Operating Characteristics curve (AUR) is a diagnostic tool for evaluating the ability of a model to rank positive instances relative to positive instances. An AUR of 50% represents a worthless prediction that performs no better than the toss of a coin. The Receiver Operating Characteristics curve plots the proportion of actual positive outcomes that are predicted positives (the true positive rate or sensitivity) against the proportion of actual negative outcomes that are false positive prediction (the false positive rate or one minus specificity) at different thresholds.

⁹ Misclassification rate at 3% is the proportion of all cases in which an error in prediction is made (either a false positive or a false negative prediction) using the top 3% of scores to define predicted positives. To allow comparison, the scored study population is the cohort born in 2007 across all models.

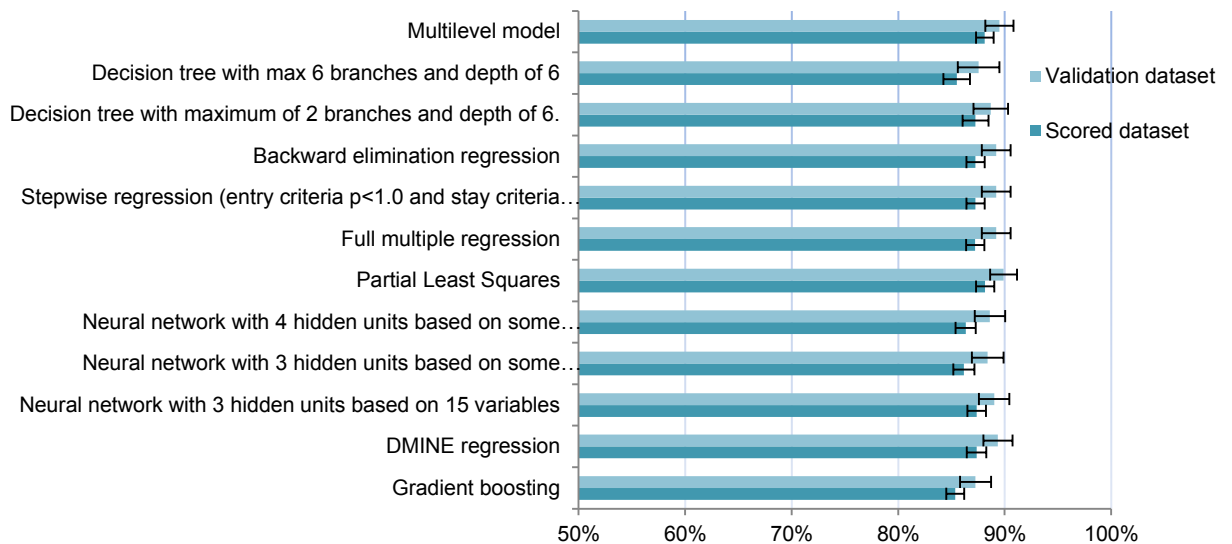
¹⁰ Cumulative Lift is a measure of model performance that is useful when looking at the most at-risk part of the population. It gives the ratio of the sensitivity of the model at the given threshold to the sensitivity resulting from a random selection of individuals from the population. Cumulative lift of 8 or 9 at 3 percent threshold indicate that if we used the models to identify the most at-risk three percent of the population, we would find eight to nine times more children with findings of maltreatment by age two than if we randomly selected three percent of children.

¹¹ Positive Predictive Value (PPV) is the proportion of children in the predicted high priority group referred who subsequently have substantiated findings of maltreatment.

¹² Sensitivity measures the proportion of children who go on to have substantiated findings of maltreatment who are correctly identified as being high priority and referred

61 Across all models the AUR for the validation dataset range from 87 to 90 percent and 85 to 88 percent for the scored cohort. The overlapping 95% confidence interval bars suggest little differences in AUR across models. Appendix 2 contains selected model diagnostics for all models developed and tested in this study.

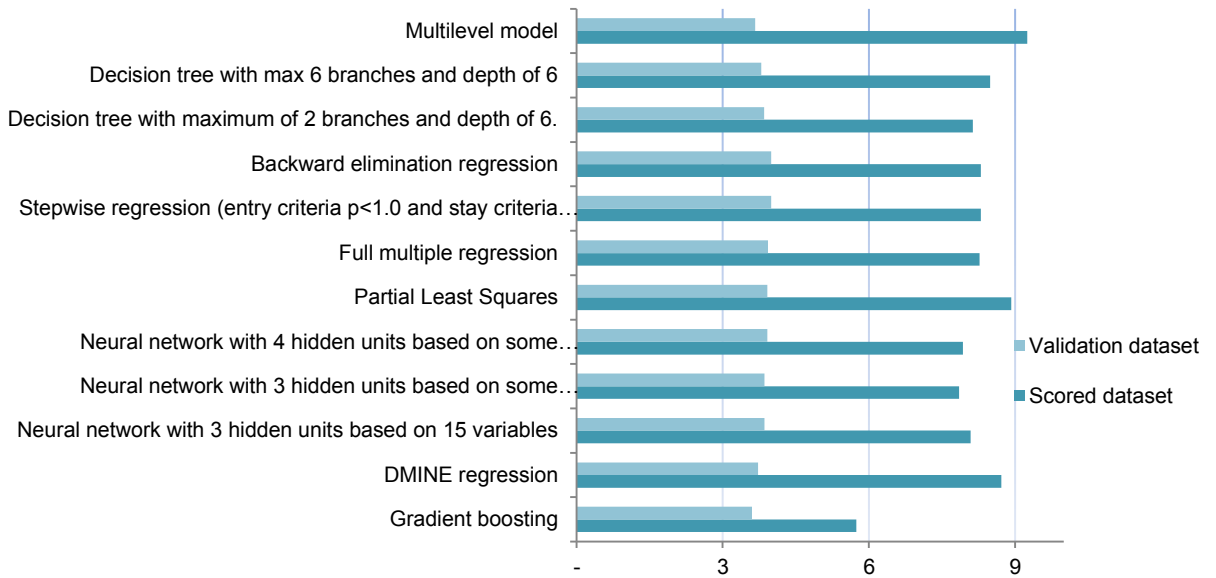
Figure 2. The AUR of various algorithms tested for the base model.



62 Cumulative lift at 3 percent threshold for validation and scored dataset is presented in Figure 3. It should be noted that cumulative lift for validation datasets range between 3 and 4 where for scored datasets it ranges from 6 to 9. This is due to the differences in the proportion with the predicted outcome in the validation and scored datasets. The rate of maltreatment findings by age 2 in the validation dataset is 20% due to the sampling method applied, whereas for scored population it is 2.3%. Nevertheless, the comparison between various algorithms across validation and scored datasets is valid.

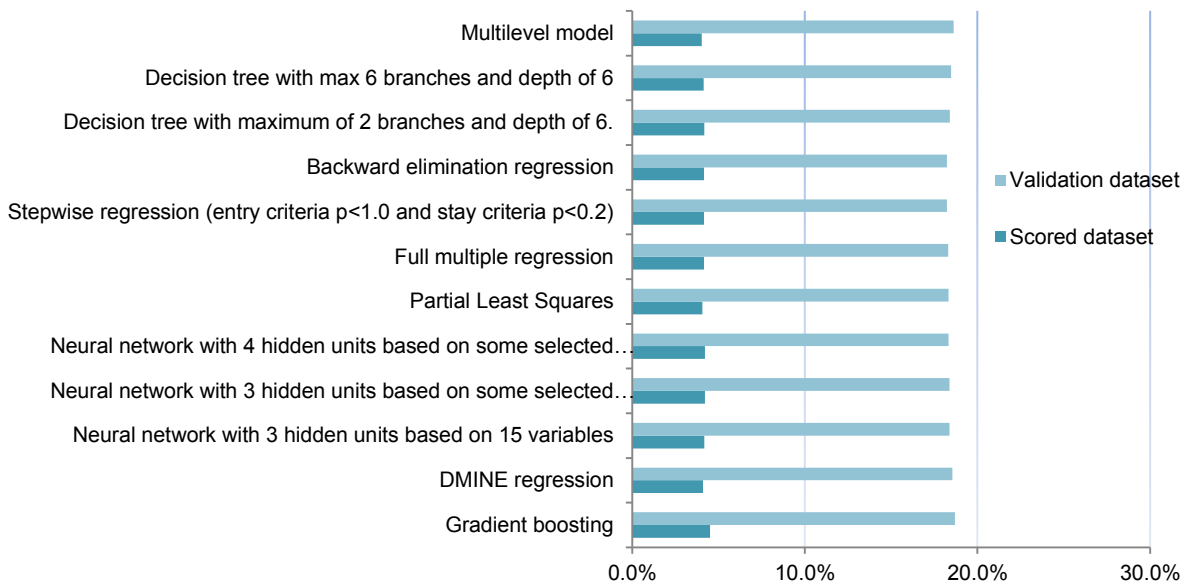
63 Once various algorithms developed on cohort 2010 are applied to cohort 2007, the Multilevel model, partial least squares model and DMINE regression models have higher cumulative lift compared to other algorithms tested, even though on validation dataset the cumulative lift across models is reasonably smaller. In a later section on the handling of local level variables we examine whether Multilevel models developed for several cohorts consistently perform better than Stepwise Logistic regression.

Figure 3. The cumulative lift of various algorithms tested for the base model.



64 Misclassification rate at 3 percent threshold on validation and scored datasets is presented in Figure 4 below. As with cumulative lift, the rate of events in the dataset (which differs between the scored and validation datasets because of the sampling method) will affect the rate of misclassification; however it is still possible to compare misclassification rate across the various algorithms. The misclassification rate across all algorithms tested is reasonably similar. At the 3% threshold level, the misclassification rate on the scored dataset is about 4% across all algorithms (see Figure 4 below).

Figure 4. The misclassification rate at 3% threshold level for various algorithms tested on the base model.



65 Stepwise logistic regression based on maximum likelihood estimator (MLE) method was selected as a modelling strategy for this feasibility study for three reasons:

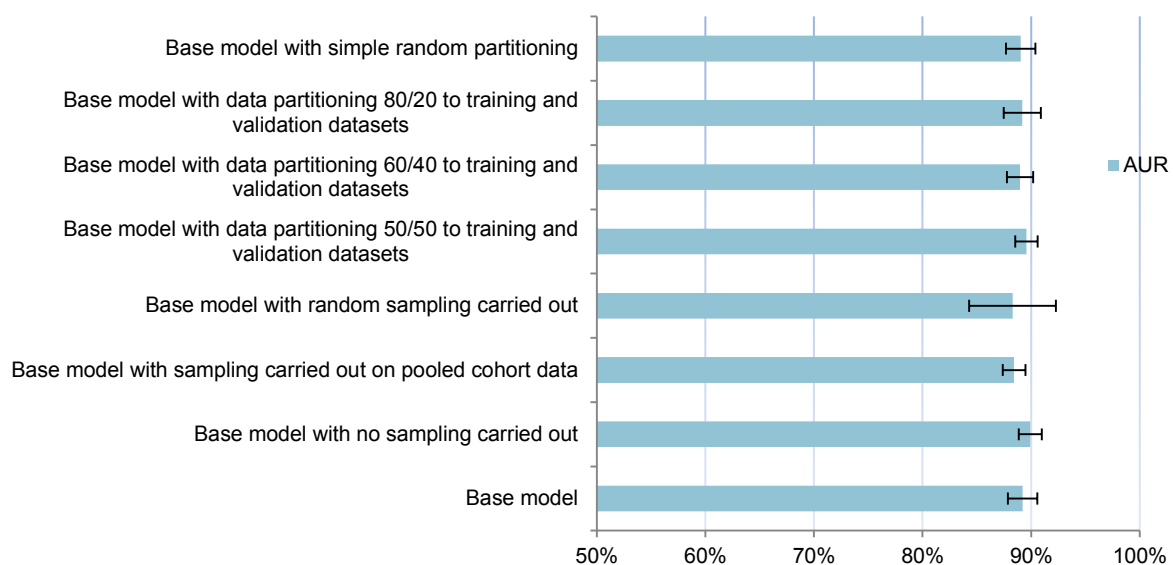
- Stepwise logistic regression has as good predictive performance as many of the other algorithms tested
- Stepwise logistic regression is transparent and can be easily explained to stakeholders, and is straightforward to implement.
- Stepwise logistic regression allows researchers to gain knowledge about the most important and useful predictors, which was a particular focus for the feasibility study.

66 Based on the results outlined above, alternative modelling algorithms perform as well and could be chosen for any future implementation.

Sensitivity to sampling and partitioning

67 We also tested whether sampling and partitioning methods will affect the performance of the base model. The results show that the sample size and partitioning method have little impact on the performance of the base model. Figure 5 presents the AUR of different models run to check for sensitivity of the base model to sampling and partitioning methods.

Figure 5. The AUR of models to test for sensitivity to sampling and partitioning (the base model).



68 Overlapping confidence interval bars suggest that in terms of predictive performance measured by AUR, models are similar. The relatively wide 95% confidence interval for the base model with random sampling suggests that sampling method has greater impact on performance of the model than the size of the modelling dataset. In order to construct the models we captured all cases of children where substantiated findings events were found, whereas random sampling randomly selects cases with events and no events.

Collinearity diagnostics

- 69 Collinearity diagnostics are conducted for predictor variables that remained in the stepwise models in order to check how predictor variables relate to each other and how this affects the stability and variance of the regression estimates.
- 70 A conditioning index from 10 to 100 indicates that collinearity affects estimates (Belsley, Kuh, and Welsch, 1980). Across all models, the conditioning index (intercept adjusted) is less than 10, with one exception. For the split model, for the “Non-Maori, and ethnic group not recorded” sub-population, the conditioning index is over 10. This suggests that there is no sign of multicollinearity problems in all models tested except the split model for the group above. In any implementation, further work could be undertaken to address this.
- 71 Variance Inflation Factor (VIF) values exceeding 5 implies that the associated regression coefficients are poorly estimated because of multicollinearity (Montgomery et.al., 2001). Across all models VIF range from 1.1 to 3.2. The model developed for the “Non-Maori, and ethnic group not recorded” group had the highest value (3.2), suggesting that the multicollinearity for that model problem did not affect regression coefficients. However, we recommend that in any implementation, variable construction and predictors be reviewed to improve the model. Table 8 present the summary of Collinearity diagnostics for all models.

Table 8. Summary of Collinearity diagnostics

Linkage	Model description	Maximum Condition index (intercept adjusted)	Maximum Variance Inflation factor (VIF)
3c	Base model for cohort 2010	6.7	2.0
Cohort sensitivity analysis			
3c	Base model for cohort 2009	7.8	3.2
	Base model for cohort 2008	1.9	1.3
	Base model for cohort 2007	1.8	1.3
PRM sensitivity to various administrative data and linkages			
1c	Model based on benefit and care protection data only	6.5	2.2
2c	Model based on benefit, care and protection data and Corrections data	6.9	2.3
4c	Model based on benefit, care and protection, Corrections and Health data	1.6	1.2
3	Model for cohort 2010 using less conservative link 3	2.0	1.4
Separate models to correct for over-representation			
	Model where other children with history of contact with CYF	1.0	1.0
	Model where no other children with history of contact with CYF	7.4	2.1
3c	Model for children of benefit caregivers based on link 3c	7.3	2.1
	Model for children of non-benefit caregivers based on link 3c	1.5	1.2
	Model for Maori ethnic group	6.1	2.0
	Model for Non-Maori and ethnic group not recorded	10.5	3.2
4c	Model for children of benefit caregivers based on link 4c	6.9	2.1
	Model for children of non-benefit caregivers based on link 4c	1.4	1.1
Modelling for various outcomes			
	Model to predict notifications (including Police FV notifications and contact records)	2.3	1.5
3c	Model to predict investigations or Child and Family Assessments	2.0	1.5
	Model to predict substantiated physical or sexual abuse or neglect (excluding emotional abuse findings)	1.7	1.2
4c	Model to predict substantiated physical or maltreatment or marker injury hospitalisation	2.0	1.5
Sensitivity to local level variables			
	Model with no local variables	2.2	1.5
3c	Model with aggregated local level variables	2.2	1.5
	Model with aggregated local level variables based on multiple counts of notifications and investigations	2.2	1.5

Performance of the base models

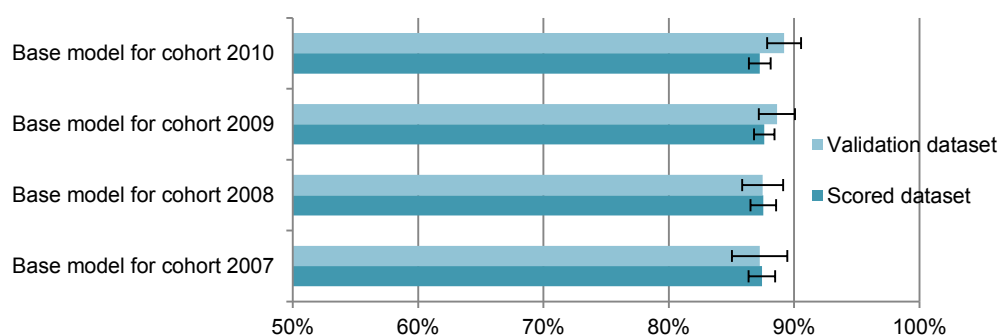
- 72 Stepwise logistic regression was selected as a modelling method where the significance entry level for predictor variables was set to $p < 1.0$ allowing all variables to enter the model. The significance stay level was set to $p < .02$, only allowing variables to remain in the model if the significance of variables is less than $p < .02$ when the effect of other variables in the model was controlled.
- 73 In the 2010 base model, 13 out of the 15 variables considered satisfy the criteria and remain in the model. Table 9 presents a summary of stepwise selection.

Table 9. Summary of stepwise selection

Step	Variable	DF	Chi-Square(at entry)	Sig.	Validation Error Rate
1	Other children with care and protection history	1	2,765.6	<.0001	4,328
2	Time on benefit in the last 5 years	3	1,087.4	<.0001	3,813
3	Caregiver with care and protection history	1	200.6	<.0001	3,698
4	CYF service centre	42	210.4	<.0001	3,717
5	Benefit caregiver is not a birth registration parent	2	72.2	<.0001	3,705
6	Mental health issues in the last 5 years	4	87.9	<.0001	3,720
7	Family violence	2	67.5	<.0001	3,710
8	Caregiver's age	5	67.2	<.0001	3,683
9	Corrections history in the last 5 years	3	45.2	<.0001	3,652
10	Single parent	2	35.8	<.0001	3,653
11	Benefit caregiver's address changes in the last year	3	33.9	<.0001	3,657
12	Findings of behavioural or relationship difficulties as a child	1	21.0	<.0001	3,643
13	Parenting demands	2	13.6	0.0011	3,629

- 74 The base model was also developed on cohorts 2007, 2008 and 2009 and applied to score the cohort 2007. The base models tested on different cohorts are comparable in terms of all model diagnostics. A comparison of AUR is presented in the Figure 6. A summary of selected model diagnostics for these models is presented in Appendix 2. Appendix 3 of this report presents the summary stepwise selection for all models tested in this study. Parameter estimates and odds ratios for all stepwise logistic regression models are provided in Appendix 4.
- 75 Across base models on different cohorts, the top predictors are reasonably stable (see Figure 4 of the main report). Overlapping confidence interval bars for AUR suggest little variation in the performance of base models developed on different cohorts.

Figure 6. The AUR of the base models developed on different cohorts and applied to cohort 2007.

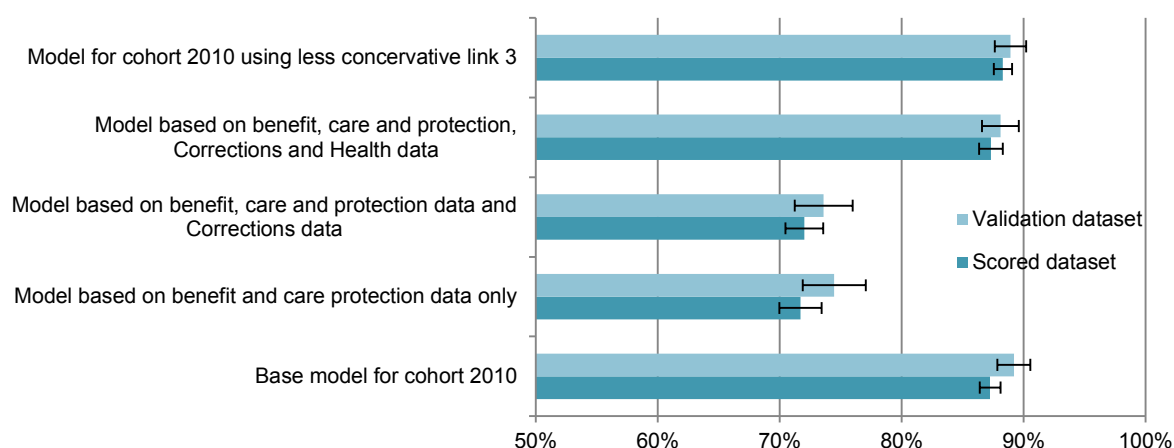


76 Other model performance indicators (misclassification rate and cumulative lift at the 3 percent threshold) performed similarly across base models for different cohorts, suggesting that the PRM model is reasonably stable and performs reasonably well despite policy changes and data changes in the past.

Performance of models on various administrative data and linkages

77 As a part of sensitivity testing, PRM models developed on various administrative data and linkages were developed. Across models, the studied populations, the predictor variables and the rate of outcomes vary. Hence, not all model fit diagnostics are easily comparable. Nevertheless, the comparison of each model's AUR for scored and validation provides an indication of the overall performance of the model and whether the models are reasonably stable. Figure 7 presents the AUR for models developed on various administrative data and linkages.

Figure 7. The AUR for models on various administrative data and linkages (validation and scored datasets)



78 The AUR of the model based on benefit, care and protection data and the model based on benefit, care and protection data and corrections data are comparable. These models are developed for the population of just children seen on benefit within 3 months of birth, while the other models also include children from birth records but never seen on benefit. In terms of AUR these models are very similar and overlapping confidence intervals suggest little differences in the AUR. The cumulative lift, misclassification rate, positive predictive value (PPV) and sensitivity measures by age 2 on the scored dataset is slightly better for the model based on benefit, care and protection data only. However the difference is very modest (see Appendix 2).

79 The model for cohort 2010 using less conservative link 3 and the model based on benefit, care and protection, correction and health data are comparable with the base model. On scored dataset, in terms of AUR, cumulative lift and misclassification rate, none of these models is consistently better than the base model on all diagnostics. However, in terms of PPV and Sensitivity by age 2 at the 3 percent threshold, the model using less conservative link 3 produces higher PPV and sensitivity (see Appendix 2).

Performance of models to correct for over-representation

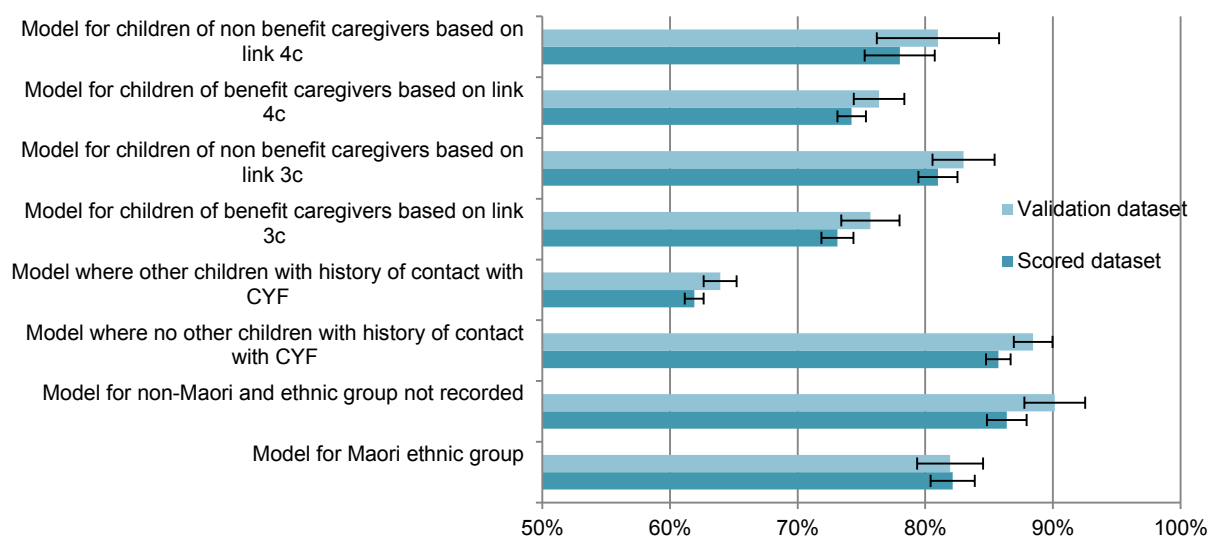
80 Other than the base model developed on various cohorts, we tested eight separate models for various sub groups of the population to correct for over-representation and

estimate whether the profile of children to be sent for referral will be similar to the profile of children with known findings of maltreatment on other dimensions. Tested separate models include the following:

- Models where other children with history or no history of contact with CYF
- Models for Maori ethnic groups and for “Non-Maori” ethnic groups
- Models for children of benefit caregivers and children of non-benefit caregivers using link 3c
- Models for children of benefit caregivers and children of non-benefit caregivers using link 4c

81 The performance of these models is not easily comparable, mainly because of differences in the studied populations and rates of maltreatment findings. It should also be noted that predictor variables used in these separate models were specifically constructed for the base (population-wide) model. It is recommended that if separate models are to be implemented, variable construction for each be reviewed. Figure 8 presents the AUR of these models and Appendix 2 contains selected model diagnostics.

Figure 8. The AUR for models to correct for over-representation (validation and scored datasets)

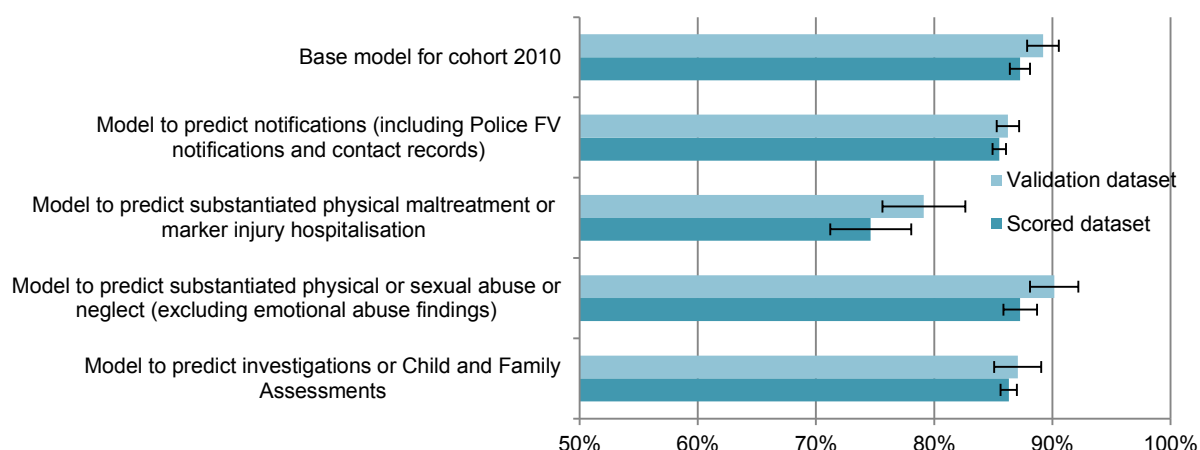


82 It is apparent that for some split models the AUR is very low. For example, the model where other children had history of contact with CYF, the AUR range between 60 to 65 percent, suggesting that this model has low accuracy in ranking within the groups identified. If separate models are to be implemented, it is recommended that further modelling examines the reclassification of existing variables and exploring the possibility of creating new set of predictor variables specifically for these groups.

Performance of models targeting different outcomes

83 Using the same predictor variables as for the base model, we assessed how well these variables can predict notifications, investigations of maltreatment and various type of maltreatment. The AUR for these models is compared with the AUR of the base model. Figure 9 presents the AUR for these models and Appendix 2 contains selected model diagnostics.

Figure 9. The AUR of the models targeting different outcomes (validation and scored datasets)



84 It is apparent that the model targeting physical abuse or maltreatment of marker injury hospitalisation is not performing as well as the base model, in terms of AUR. Some selected model diagnostics are presented in the Appendix 2.

Handling local-level variables

85 Peer review of preliminary findings raised concerns that the models over-emphasise individual and family related risk factors and that risks associated with the environment and administrative context are not well captured. The suggestion was to adopt a multilevel framework where predicted estimates of person-level risk will be significantly more realistic. Following this recommendation, we conducted several sensitivity tests with local-level variables.

86 In the base models, a variable that indicates which CYF Service Centre the child would be served by given their place of residence was newly included in an effort to account for a number of dimensions on which local conditions might impact on outcomes for a child. These include, for example:

- the level of deprivation in the community and locality-specific relationships between alcohol outlet density and alcohol-related social harms
- social norms, the availability and co-ordination of services, and the degree to which there is a sense of collective responsibility for positive child development in the community
- organisational and case worker factors that can vary across local care and protection service offices, an important source of variation in decision making and recorded substantiations of maltreatment.

87 When the CYF service centres are included in the model, comparison is made between children with similar characteristics within same CYF service centres.

88 We also tested series of multilevel models on different cohorts, compared the performance of models and looked at the profile of the 3000 children with highest PRM scores. This analysis is presented in the following section of this report.

89 We also created a number of other local-level variables designed to proxy community deprivation, community vigilance and the rate of investigation in the community, where

community is defined by boundaries of CYF service centre (see Appendix 1 for detailed description). We tested several versions of the model with local-level variables defined slightly differently and compared them with the models with no local level variables and the base model. Main findings of the analysis are presented below.

Multilevel model versus Stepwise Logistic regression

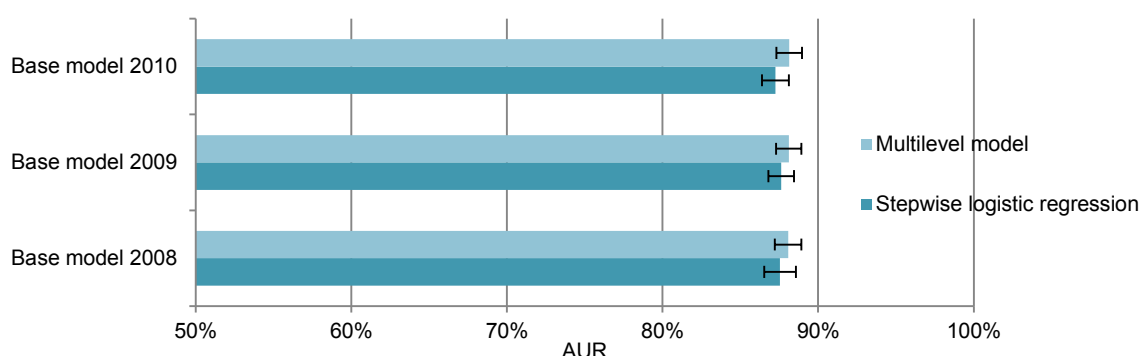
90 Multilevel modelling or Hierarchical Linear Model (HLM) is becoming a popular modelling methodology in the field of social science. There are two known advantages of using HLM over MLE method, which is an underlying method for Stepwise Logistic Regression. These are:

- multilevel modelling takes into account the hierarchical structure of the data when estimating standard errors. It considers the correlated errors between levels and provides more realistic and conservative statistical testing (Ferron et.al, 2004) (in this study the level 1 is individual/family level and level 2 is community level, defined by CYF service centres).
- multilevel modelling efficiently deals with interactions (Ferron et.al, 2004).

91 Multilevel modelling was considered as an alternative method for the base model. The results indicated that the multilevel model performs slightly better than stepwise logistic regression (see Figure 2 and Figure 3 above). Here, we run a series of multilevel models on different cohorts to check whether these findings are consistent. Multilevel models were developed on several cohorts and compared against the same models developed using Stepwise Logistic regression based on MLE. Figures 10 presents the AUR from these models and Appendix 2 contains other model diagnostics.

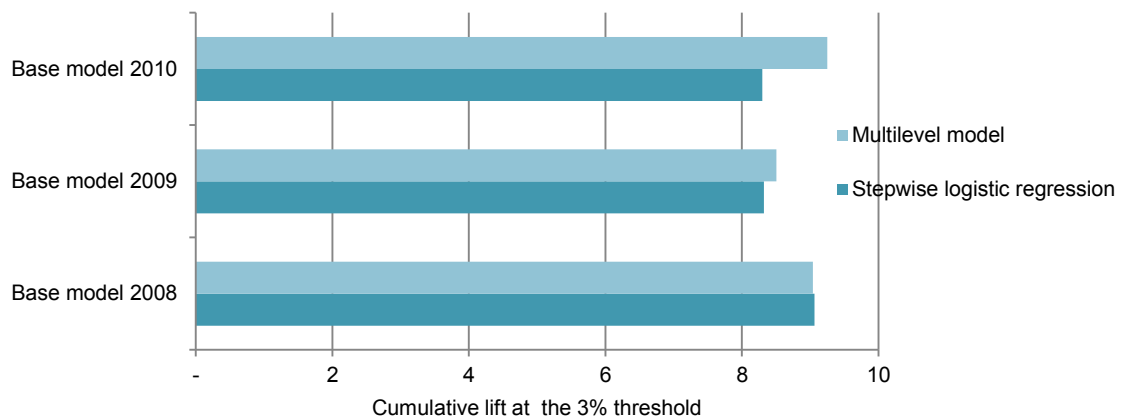
92 In terms of AUR, multilevel models produce slightly higher AUR. However, overlapping confidence intervals suggest little differences in AUR (see Figure 10). In terms of cumulative lift, the results are not consistent. In the model developed on cohort 2008 and applied to cohort 2007, cumulative lift at the 3% threshold is slightly lower for the multilevel model than for the stepwise logistic regression (see Figure 11). Overall, the misclassification rate at the 3% threshold level from two methods is very similar across all three cohorts (see Figure 12).

Figure 10. The AUR comparison between HLM and MLE methods for base models (scored population).



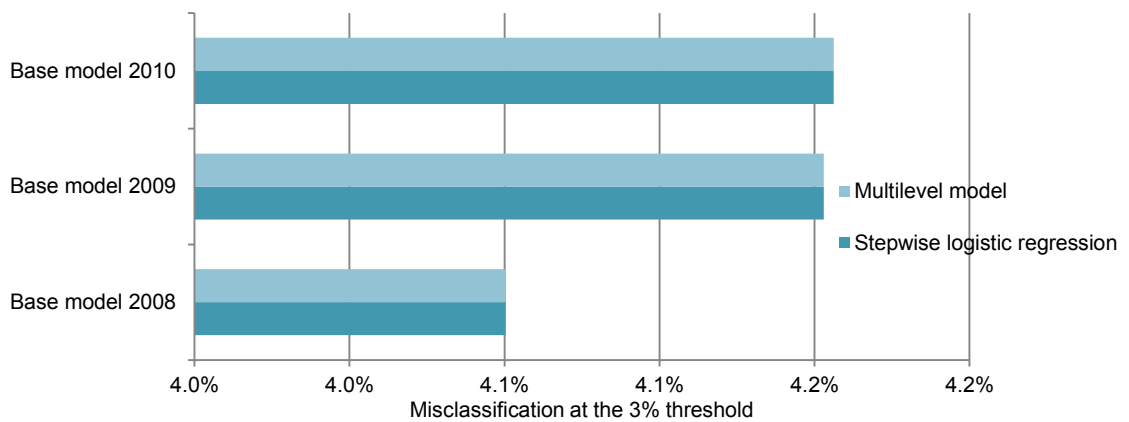
Notes: Model developed on cohort 2008 to 2010 and applied to cohort 2007

Figure 11. The Cumulative lift comparison between HLM and MLE methods for base models (scored population).



Notes: Model developed on cohort 2008 to 2010 and applied to cohort 2007

Figure 12. The comparison of misclassification rates between HLM and MLE methods for base models (scored population).



Notes: Model developed on cohort 2008 to 2010 and applied to cohort 2007

93 Overall, we conclude that when the performance of the model is measured in terms of AUR, cumulative lift and misclassification rate, multilevel modelling does perform slightly better. Especially, there are some modest gains from multilevel modelling in terms of PPV and sensitivity. However the results show that the gain is not consistent across models for different cohorts. Table 10 presents a comparison of PPV and sensitivity by age 2 from the two modelling approaches. Note that PPV and sensitivity sit at different levels comparing the scored cohort and the training and validation samples. This is due to the differences in the proportion with the predicted outcome in the training/validation and scored datasets. The rate of maltreatment findings by age 2 in the training/validation datasets is 20% due to the sampling method applied, whereas for scored population it is 2.3%. Nevertheless, the comparison of the two modelling approaches is valid.

Table 10. Comparison of two modeling approaches: PPV and Sensitivity by age 2

Dataset		PPV by age 2		Sensitivity by age 2	
		Stepwise	Multilevel	Stepwise	Multilevel
Scored cohort 2007	Base model for cohort 2008	21	21	27	27
	Base model for cohort 2009	19	20	25	26
	Base model for cohort 2010	19	21	25	28
Training	Base model for cohort 2008	85	85	13	13
	Base model for cohort 2009	88	82	13	12
	Base model for cohort 2010	86	84	13	13
Validation	Base model for cohort 2008	73	78	11	12
	Base model for cohort 2009	81	81	12	12
	Base model for cohort 2010	80	73	12	11

94 Offsetting the small gains in performance measures, multilevel models tended to increase the over representation of sub groups within the group with highest risk scores. When the profile of top 3000 children of cohort 2007 scored using the two methods is compared, the profile of children with the highest PRM scores from a multilevel model is broadly similar to the profile of children with the highest scores using Stepwise Logistic Regression. However, it is notable that the multilevel model increased over-representation relative to known maltreatment on some dimensions. Table 11 presents the profiles of 3000 children with highest PRM scores from both models.

Table 11. Profile of children of cohort 2007 with findings of maltreatment and 3000 children with the highest scores from the 2010 multilevel and Stepwise Logistic Regression models

	% of children with findings by age 2	% of the 3,000 children with the highest PRM scores (multilevel model)	% of the 3,000 children with the highest PRM scores (Stepwise logistic regression)
Māori child (on non-missing)	60.9	69.1	69.2
Male child	53.0	54.1	52.5
Low birth weight child or pre-term birth	10.7	9.6	8.2
Multiple birth child, other children under 2 or 3+ children	32.9	32.9	38.1
No other children (estimated) ⁽²⁾	51.1	51.1	45.7
Other children with a care and protection history in the last 5 years ⁽¹⁾	34.9	61.3	59.9
Other children with a Police family violence notification or contact record in the last year	8.6	12.8	12.4
Single parent	74.3	89.1	87.7
No birth registration at 3 months of age	26.1	26.1	35.5
Mother or caregiver aged under 25	53.5	58.7	54.6
Parents or caregivers with a care and protection history as a child	43.2	63.5	57.2
Parents or caregivers with findings of behavioural or relationship difficulties as a child	19.1	32.1	29.1
Child seen on benefit by 3 months of age	79.0	93.3	92.3
Mother or caregiver's spent 80-100% of time on benefit in the last 5 years	50.7	68.8	65.5
Parents or caregivers received benefit for a substance abuse disorder in the last 5 years	8.2	14.1	14.1
Parents or caregivers received benefit for other mental health disorder in the last 5 years	17.4	19.0	19.0
Primary caregiver's with 3+ addresses recorded in benefit data in the last year	9.9	15.4	15.4
Parents or caregivers with any sentence in the last 5 years	25.5	25.5	33.6
Parents or caregivers with a custodial sentence in the last 5 years	11.1	11.1	16.6
High deprivation neighborhood (deciles 8-10, of non-missing)	69.0	70.7	70.7

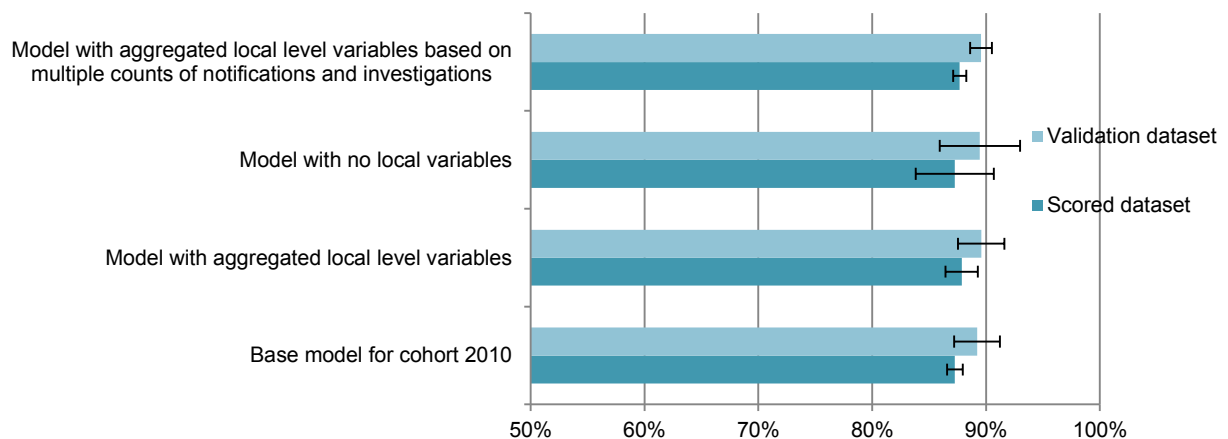
Notes: (1) Excludes other children with Police family violence notifications but no other history.

(2) If children are identified through birth data only, this indicator is set to one if the birth registration indicates that there have been no previous children from the same parent relationship and there are no New Zealand registered previous births to the same mother (looking back to 2004). Note that this will overstate cases where there are no other children where it misses previous children from a different parent relationship and where it misses previous New Zealand registered births that occurred prior to 2004.

Performance of various models with local level variables

- 95 In addition to series of multilevel models, we tested the sensitivity of the base model to alternative specifications of local-level variables. The alternative variables included aggregated proxies for community deprivation, vigilance and the rate of investigation, with tests of sensitivity to whether measures of community vigilance and the rate of investigations performed better when calculated taking account multiple counts of events (“model with aggregated local-level variables based on multiple counts of notifications and investigations”), rather than a count of children with events (“model with aggregated local-level variables”).
- 96 Out of three alternative local-level variables tested, only the rate of investigations remained in the models after the stepwise selection process. Both community vigilance and community deprivation variables were dropped from the models. The AUR of the various models is presented in Figure 13.

Figure 13. The AUR comparison of models with various local level variables



- 97 The AUR of the base model and the model with no local variables is almost the same. However the 95% confidence interval for AUR is much smaller on base model. Table 12 summarises the model fit statistics for various models and present PPV and Sensitivity by age 5 at the 3% threshold of the model once model developed on cohort 2010 is applied to cohort 2007.

Table 12. The performance of the models with local level variables (scored population)

	Base model for cohort 2010	Model with no local variables	Model with aggregated local level variables	Model with aggregated local level variables based on multiple counts(1)
Cumulative lift at 3 %	8.3	8.9	8.3	8.6
Misclassification rate 3% threshold level	4.2	4.1	4.2	4.1
PPV by age 5	33.9	35.4	33.9	35.2
Sensitivity by age 5	21.1	22.0	21.1	21.9

Notes: (1) The local level variables are based on multiple counts of notifications and investigations. Models are developed on cohort 2010 and applied to cohort 2007

98 Overall, the results suggest that, for the purposes of prediction, stepwise regression results have little sensitivity to the handling of local-level variables, or their inclusion or exclusion. These findings are specific to our examination of variables aggregated across CYF sites. In any implementation, sensitivity to aggregation of different variables at different levels (eg. immediate neighbourhood rather than CYF site) could be explored further.

Profile of scored cohort 2007

99 Based on the parameter estimates of the base model developed on cohort 2010, the population of birth cohorts 2007 to 2010 is scored in terms of risk of maltreatment. Appendix 5 of this report presents the descriptive profile of all children of cohort 2007, compared with population of cohort 2007 with outcomes and the 3000 children with the highest scores across various models. It should be noted that the number of children with highest PRM scores from the models to correct for disproportionality are based on their known share of maltreatment.

100 For the base model, Figure 14 and Figure 15 show graphically the distribution of probabilities by the outcome targeted (findings of maltreatment by age 2) and by findings of maltreatment by age 5.

Figure 14. Predicted probabilities of birth cohort 2007 by outcome targeted (findings of maltreatment by age 2), using algorithm developed on cohort 2010.

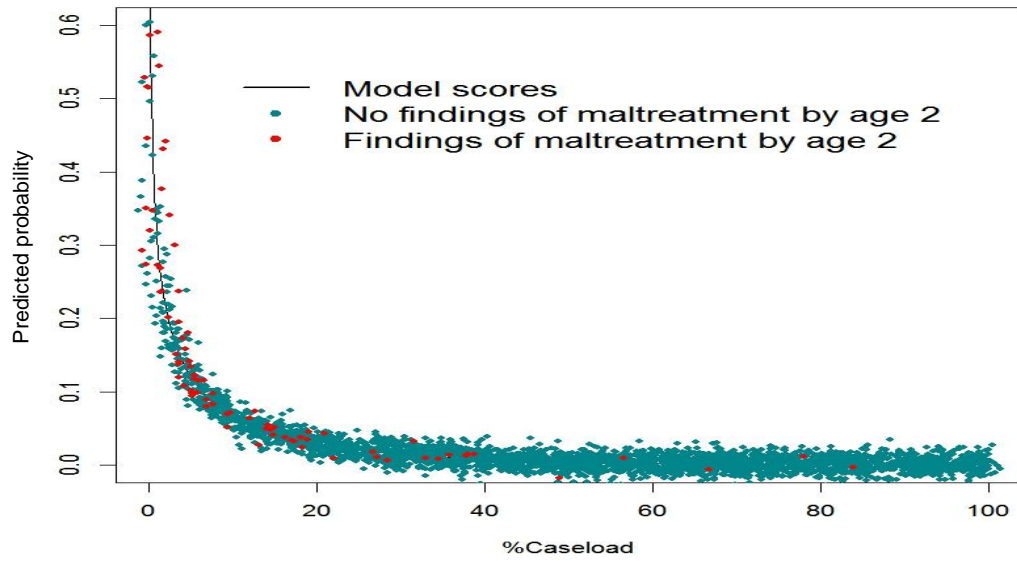
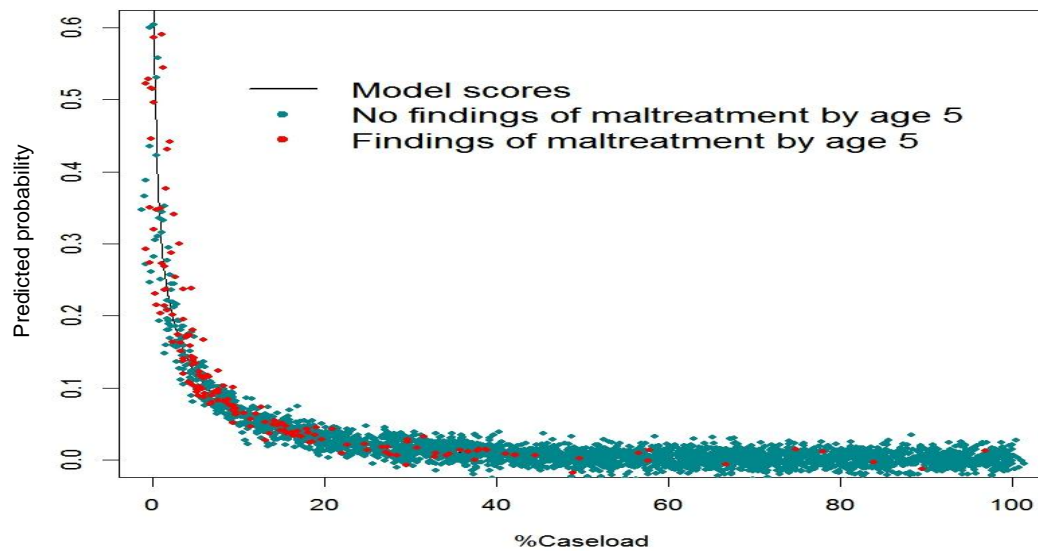


Figure 15. Predicted probabilities of birth cohort 2007 by findings of maltreatment by age 5, using algorithm developed on cohort 2010.



Appendix 1 – Predictor variable derivation

Gender of child

Values: 1-male, 2-female

For all links, if children are identified through benefit data only, gender is as indicated in benefit data.

For links 3c and 4c, if children are identified through birth data only, gender is as indicated on the birth registration.

For links 3c and 4c, if children are identified through birth data and benefit data, birth registration information is prioritized.

In the small number of cases where gender is not recorded, it is imputed using the child's first name.

Low birth weight or pre-term

Values: 1-yes, 2-no / unknown

Low birth weight is defined as birth weight up to 2,500 grams. Pre-term is defined as gestation less than 37 weeks.

For link 3c, this information is from birth data. It is sourced from health provider birth notifications sent to the Department of Internal Affairs (it is not reported by parents on the birth registration form).

For link 4c, it also draws on Health data for live births in public hospitals (plus publicly-funded births in private hospitals and primary birthing units).

For links 3c and 4c, if children are identified through benefit data only, birth weight and gestation is missing and the variable is coded 2-no/unknown.

Parenting demands

Values: 1-high parenting demands, 2-no other children, 3-other children but not high demands

For all links, if children are identified through benefit data only, this variable is set to 1 (high parenting demands) if:

- the child is a multiple birth child (i.e there is one or more other children born within a day of the reference child are included in benefit on the day the reference child is first included in benefit)
- one or more other child aged under two and at least two days older than the reference child is included in benefit on the day the reference child is first included in benefit
- three or more other children are included in benefit on the day the reference child is first included in benefit (the count of other children includes multiple birth siblings).

For links 3c and 4c, if children are identified through birth data only, this variable is set to 1 if:

- the birth registration indicates that the child is a multiple birth child
- there is another New Zealand registered birth to the same mother where the child is aged under two, and has not died since
- three or more other live children born to the same parent relationship are listed on the birth registration. The count of children born to the same parent relationship is used as a proxy for family size as birth data does not provide a count of other children being cared for by the parents.

For links 3c and 4c, if children are identified through birth data and benefit data, benefit data are prioritised as these provide better information on children being cared for.

For link 4c, the variable is also set to 1 if:

- the child has a birth abnormality recorded at birth (defined in accordance with definitions used in the New Zealand Birth Defects Registry¹³)
- the estimated number of previous births to the mother based on an administrative count of previous births recorded in Health maternity data (looking back to 1988) is three or more.

Other children with a care and protection history in the last 5 years

Short name: Other children with care and protection history

Values: 1=yes, 2=no (base model)

This variable summarises the Child Youth and Family (CYF) care and protection history of other children of the parents or caregivers recorded in the five years prior to the date of the reference child's birth registration or first date of benefit receipt.

For all links, if children are identified through benefit data only, this history relates to all children either currently or previously included in benefit with the primary benefit recipient or their partner (where the events for children previously included in benefit occurred while those children were included in benefit with the primary benefit recipient or their partner).

For links 3c and 4c, if children are identified through birth data only, this history relates to all previous children with a New Zealand registered birth (looking back to mid-2004) to the same mother or father as the reference child.

For links 3c and 4c, if children are identified through birth data and benefit data, this history relates both to children either currently or previously included in benefit, and to previous children born to the same mother or father.

For links 3c and 4c, the variable is binary and distinguishes whether the history of the other children includes:

- no care and protection contact, or Police family violence notifications or contact records but no other notifications and no other contact
- notifications other than family violence notifications, investigations, substantiated findings of maltreatment or behavioural or relationship difficulties or placement or entry into the care of the Chief Executive.

For links 1c and 2c, the variable is categorical and distinguishes whether the history of the other children includes:

- no care and protection contact, or Police family violence notifications or contact records but no other notifications and no other contact
- notifications other than family violence notifications but no substantiated findings of maltreatment or behavioural or relationship difficulties, placement, or entry into the care of the Chief Executive
- substantiated findings of maltreatment or behavioural or relationship difficulties, placement or entry into the care of the Chief Executive.

Note that in some cases an older child's care and protection history will not be comprehensively captured by the administrative data available in electronic form (see Box 1 below for further details).

Family violence

Values: 1-events in one of the last 12 months, 2-events in more than one of the last 12 months, 3-no events (base model)

This variable captures the number of months in the year prior to the date of the reference child's birth registration or first date of benefit receipt in which Police family violence notifications or contact records

¹³ See <http://nzbd.ac.nz/assets/FILES/Final%20published%20table%202000-2009.pdf>

were received in respect of other children of the parents or caregivers.

For all links, if children are identified through benefit data only, this history relates to all children currently included in benefit with the reference child.

For links 3c and 4c, if children are identified through birth data only, this history relates to all previous children with a New Zealand registered birth (looking back to mid-2004) to the same mother or father as the reference child.

For links 3c and 4c, if children are identified through birth data and benefit data, data for other children included in benefit is used as benefit data provide better information on children being cared for.

Note that in base models, this variable captures the intensity of Police FV events only for children already known to Child Youth and Family. This is because under the conservative data linkage rules, records for children in respect of whom Police family violence contact records are received but who have had no other engagement with Child Youth and Family are not able to be linked because parent or caregiver information is not available to confirm the linkage (details of adults present at the incident are available, but the relationships of these adults to the children are not known). Under the current linkage, this prevents examination of the role of Police family violence contact records in the absence of other engagement as a predictor.

For link 4c, this variable is also set to 2 if the mother was hospitalized at least once for assault in the previous five years, or another child in the family was hospitalized for a maltreatment related injury, an intracranial injury or a long bone fracture (marker injuries that may indicate maltreatment) before age one. In both cases, short stay emergency department admissions are excluded. Categories are 1- / notification or contact record in 1 of the last 12 months / notification or contact record in 2+ of the last 12 months, or mother hospitalised for assault in the last 5 years, or other child hospitalised for maltreatment, intracranial or long-bone fracture injury by age 1 / no events.

Single parent

Values: 1-single parent, 2-single parent and no father listed on birth registration, 3-not a single parent or partnership status unknown

For all links, if children are identified through benefit data only, this indicator is set to one if there is no partner recorded as being attached to the benefit family and 3 otherwise.

For links 3c and 4c, if children are identified through birth data only, this indicator is set to one if the birth registration indicates that at the time of the child's birth, the parents of the child were not married or in a civil union or defacto relationship,¹⁴ 2 if no father is listed on the birth registration, and 3 otherwise.

For links 3c and 4c, if children are identified through benefit and birth data, this indicator is set to one if benefit data indicates single parent status, 2 if no father is listed on the birth registration, and 3 otherwise.

¹⁴ Note that where the parents separate and re-partner by the time of the child's birth, relationship information recorded on the birth certificate will be an imperfect indicator of single parent status.

Box 1 - Electronic records may not capture all of a parent, caregiver or older child's care and protection history

Electronic collection of care and protection data began in 1996. Migration from paper files to the initial electronic system, and the subsequent migration to the new CYRAS system in 2000 mark points where incomplete transfer of records may have occurred.

In some cases with the older records used in this analysis, early records about parenting issues were in some cases located on the record of an older sibling. Current practice ensures that information about a child lodged on a sibling's record gets carried over to their own record.

In some cases with the older records used in this analysis, support was provided to the baby of a young woman in care, but the child did not obtain a separate record from his or her mother until some time after birth, when a new intervention was required. Current practice ensures that information about a child lodged on caregiver's record gets carried over to their own record.

In contrast to the benefit system where the child's identity is verified with reference to a birth certificate, care and protection identities are generally established without verification, based on information provided by the parents, caregiver, or a notifier. Where a contact occurs as a result of a third party notification that was not investigated, name and date of birth information can be inherently vague (eg. where a community member makes a notification about a child and knows their first name but not their surname or date of birth). This may result in the individual's case history being spread across multiple child records.

There are a number of situations where care and protection matters may be managed by parties other than Child, Youth and Family. Under the Children, Young Persons, and Their Families Act 1989, care and protection matters may be managed within community organisations either in formal partnerships with CYF through a Court Plan, informal agreements such as family/whānau agreements or independently and autonomously through powers vested in organisations under sections 396 and 401. Section 19 referrals represent a situation where an organisation works with a family/whānau without reference to CYF and then makes a referral for a family group conference to use the statutory process to further consider the care and protection concerns. Records of the early assessment of care and protection concerns prior to the family group conference are not available as these lie with the organisation concerned. Under the provisions of the Family Court family/whānau can formally arrange for substitute care for members or arrange this without court involvement (for example, whāngai of Māori children).

In these situations, the electronic administrative data used for this study may provide an incomplete account of the care and protection history of the child.

At least one caregiver is not a birth registration parent

Values: 1=yes, 2=no birth registration, 3=no

For links 3c and 4c, if children are identified through birth data and benefit data, this variable is set to one if at least one of the caregivers is not a parent listed on the birth registration and three otherwise.

If children are identified through benefit data only (no birth registration by three months) the variable is set to 2.

If children are identified through birth data only, the variable is set to 3.

Age of mother or primary caregiver when child was born

Short name: Caregiver's age

Values: 1-under 20, 2- 20-24, 3- 25-29, 99- 30-34 or missing, 4- 35-39, 5 40 plus

For all links, if children are identified through benefit data only, this variable records the age group of the primary benefit recipient at the date of the child's birth.

For links 3c and 4c, if children are identified through birth data only, this variable records the age group

of the mother listed on the birth registration at the date of the child's birth.

For links 3c and 4c, if children are identified through birth data and benefit data, this variable records the age group of the mother at the date of the child's birth, except in cases where the child is being cared for by caregiver/s who do not include the birth registration mother, in which case the age group of the primary benefit recipient at the date of the child's birth is used.

Primary caregiver's highest educational qualification

Values: 1-qualifications missing or not recorded, 2-no qualification, 3-degree or professional qualification, 4-post-school qualification 5-school qualification

This categorical variable is only considered for links 1c and 2c due to large numbers of missing values, even where the child is identified through benefit data. It indicates the highest educational qualification of the primary benefit recipient recorded in benefit data.

Note that educational qualifications are not recorded for all benefit recipients, as capture of this information depends on work obligations and case management practices which vary across benefit types and over time.

Parents or caregivers with a care and protection history as a child

Short name: Care and protection history as a child

Values: 1-yes, 2-no (base model)

This variable summarises the care and protection history of the parents or caregivers in their own childhood and adolescence.

For all links, if children are identified through benefit data only, this history relates to the caregivers (primary benefit recipient and their partner).

For links 3c and 4c, if children are identified through birth data only, this history relates to the parents recorded in birth data.

For links 3c and 4c, if children are identified through birth data and benefit data, this history relates to the parents recorded in birth data, except in cases where the child is being cared for by caregiver/s who do not include the birth registration mother, in which case this history relates to the caregivers.

For links 1c and 2c, the variable is categorical and distinguishes whether the caregivers' childhood care and protection history includes:

- no care and protection contact
- notifications but no substantiated findings of maltreatment, placement, or entry into the care of the Chief Executive
- substantiated findings of maltreatment, placement or entry into the care of the Chief Executive.

For links 3c and 4c, the variable is binary and distinguishes whether parents' or caregivers' childhood care and protection history includes:

- no care and protection contact
- notifications, investigations, substantiated findings of maltreatment, or placement or entry into the care of the Chief Executive.

Note that in some cases a parent or caregiver's care and protection history will not be comprehensively captured by the administrative data available in electronic form (see Box 1). This is more likely to be the case for older parents and caregivers as electronic capture of care and protection events did not begin until 1996.

Parents or caregivers with findings of behavioural or relationship difficulties as a child

Short name: Behavioural or relationship difficulties as a child

Values: 1=yes, 2=no

This variable indicates whether the parents or caregivers had substantiated findings of behavioural or relationship difficulties in childhood or adolescence.

For all links, if children are identified through benefit data only, this indicator relates to the caregivers (primary benefit recipient and their partner).

For links 3c and 4c, if children are identified through birth data only, this indicator relates to the parents recorded in birth data.

For links 3c and 4c, if children are identified through birth data and benefit data, this indicator relates to the parents recorded in birth data, except in cases where the child is being cared for by caregiver/s who do not include the birth registration mother, in which case this indicator relates to the caregivers.

Note that in some cases a parent or caregiver's findings of behavioural or relationship difficulties in childhood will not be comprehensively captured by the administrative data available in electronic form (see Box 1). This is more likely to be the case for older parents and caregivers as electronic capture of care and protection events did not begin until 1996.

Mother or primary caregiver's time on benefit in the last 5 years

Short name: Time on benefit in the last 5 years

Values: 1- more than 80%, 2- 20<-80%, 3-up to 20%, 4-no time

This categorical variable indicates the proportion of time the parent or caregiver spent supported by main benefits (Box 2) in the previous five years. For younger parents and caregivers, it includes time spent supported by benefit as a dependent child.

For all links, if children are identified through benefit data only, this indicator relates to the primary benefit recipient.

For links 3c and 4c, if children are identified through birth data only, this indicator relates to the mother recorded in birth data.

For links 3c and 4c, if children are identified through birth data and benefit data, this indicator relates to the mother recorded in birth data, except in cases where the child is being cared for by caregiver/s who do not include the birth registration mother, in which case this indicator relates to the primary benefit recipient.

Note that while this variable is used in this analysis as a proxy for poverty persistence, parents and caregivers who spend long periods on low income but do not receive benefits are not identified. A possible enhancement would be to include time spent receiving Accommodation Supplement which tapers out further up the income distribution but can be used as a proxy for low income among those not receiving main benefits.

Box 2 - Benefit types

New Zealand has a non-contributory welfare benefit system made up of three distinct tiers of provision: main benefits; supplementary assistance payments and tax credits. This study focuses on children identified through inclusion in main benefits. Main benefits most commonly received by parents and caregivers of children over the period covered by this study include:

- unemployment and training related benefits (paid where a person was seeking full-time work or in approved training aimed at helping the person to find work)
- Domestic Purposes Benefit – Sole Parent and Emergency Maintenance Allowance payable to sole parents
- Sickness Benefit (paid to people who cannot work or work reduced hours due to sickness injury, disability or pregnancy). Sickness Benefit paid in the latter weeks of pregnancy can continue to be paid for up to 13 weeks after the birth of the child. In most cases where newborn children are supported by Sickness Benefit in their first weeks of life, the mother will then transfer to Domestic Purposes Benefit – Sole Parent)
- Invalid's Benefit (paid to people with a long-term and severe incapacity).

Other main benefits received less frequently by parents and caregivers include Emergency Benefit, Domestic Purposes Benefit for carers and women alone, Widow's Benefit, New Zealand Superannuation (a non-contributory retirement pension), Orphan's Benefit (payable where the child's parents are deceased) and Unsupported Child's Benefit (payable to the caregiver of a child whose parents can't support them as a result of family breakdown or removal from their parents' care).

Main benefits are generally subject to a test of the joint income of the primary benefit recipient and their partner; the benefit reduces as joint private income increases. The exceptions are New Zealand Superannuation, Orphan's Benefit and Unsupported Child's Benefit which are non-income tested.

There is no test of assets, except in the case of benefits such as Emergency Benefit and Emergency Maintenance Allowance which are paid on the grounds of hardship to clients who would do not meet the conditions of entitlement for other main benefits.

Main benefits can be paid together with:

- supplementary benefits (payable to people on low and middle incomes, including people not receiving main benefits, to help with a specific need or specific cost – examples include Accommodation Supplement and Disability Allowance)
- family tax credits (payable to low and middle income families with dependent children, including families not receiving main benefits). (Ministry of Social Development, 2012)

Parents or caregivers' known mental health issues in the last 5 years

Short name: Mental health

Values: 1-substance abuse issues, 2-persistent substance abuse issues, 3-mental health issues other than substance abuse, 4-persistent mental health issues other than substance abuse 5-no known mental health or substance abuse issues

This categorical variable summarises indicators of mental health issues of the parents or caregivers that can be observed in administrative data. Substance abuse is prioritised so that if at least one of the parents or caregivers has an indicator of substance abuse and they or the other parent has other mental health disorders, the child is categorized as having a parent or caregiver with substance abuse issues.

Persistent substance abuse or other mental health issues are defined as those indicated in at least three of the last 5 years. Persistence is assessed for each parent or caregiver individually, and the maximum value looking across parents or caregivers individually is used

For link 1c and 2c, indicators of mental health issues in each of the previous five years are based on incapacity codes recorded if the caregivers (primary benefit recipient and their partner) received

Sickness or Invalid's Benefit.

For link 3c and 4c, indicators of mental health issues in each of the previous five years are based on incapacity codes recorded if the caregivers *or parents listed on the birth registration* received Sickness or Invalid's Benefit, and:

- If children are identified through benefit data only, this indicator relates to the caregivers.
- If children are identified through birth data only, this indicator relates to the parents recorded in birth data.
- If children are identified through birth data and benefit data, this indicator relates to the parents recorded in birth data, except in cases where the child is being cared for by caregiver/s who do not include the birth registration mother, in which case this indicator relates to the caregivers.

For link 4c, indicators of mental health issues in each of the previous five years also draw on the following Health administrative information for the birth mother¹⁵:

- hospitalisation discharge events in each of the previous five years (excluding short stay emergency department events) for any reason where a clinical code indicates mental and behavioural disorders due to psychoactive substance use (excluding amnesia and acute intoxication)
- hospitalisation discharge events in each of the previous five years (excluding short stay emergency department events) for any reason where a clinical code indicates other mental and behavioural disorders (including intellectual disability)
- use of community-based addiction services in each of the previous five years (including only service usage that involves face-to-face contact)
- use of mental health services that involve face-to-face contact in the previous five years (excluding addiction services). The majority of these contacts were provided in a community setting.
- prescribed pharmaceuticals in each of the previous five years used solely to treat substance abuse mental health disorders
- prescribed pharmaceuticals in each of the previous five years used solely to treat mood, or other mental health disorders (excluding substance abuse) (prescribed pharmaceuticals used solely to treat anxiety disorder were considered for inclusion but had no statistically significant bivariate association with findings of maltreatment so were excluded).

Note this variable will not provide comprehensive capture of the existence and persistence of disorder. For example, incapacity codes are only recorded for benefits for which a doctor's certification of incapacity is a condition of entitlement for benefit. As a result, the variable does not provide a comprehensive count of children with caregivers with a history of mental health disorder (certification of incapacity is not, for example, a condition of benefit receipt for Domestic Purposes Benefit – Sole Parent). In some cases, the variable may overstate the presence of disorder. For example, pharmaceuticals may be prescribed in a non-standard way to individuals without mental health disorders.

Mother or primary caregiver's address changes in the last year

Short name: Address changes in the last year

Values: 1-no address changes, 2- 1 or 2 address changes, 3- 3 plus address changes, 4- missing

This categorical variable provides a proxy for the number of changes in residential address recorded.

For all links, if children are identified through benefit data only, this indicator relates to the primary benefit recipient.

For links 3c and 4c, if children are identified through birth data only, this indicator relates to the mother

¹⁵ Identified in Health data if she is one of the caregivers or parents listed on the birth registration (depending on whose data is being used).

recorded in birth data.

For links 3c and 4c, if children are identified through birth data and benefit data, this indicator relates to the mother recorded in birth data, except in cases where the child is being cared for by caregiver/s who do not include the birth registration mother, in which case this indicator relates to the primary benefit recipient.

For links 1c, 2c and the variable counts changes in residential addresses recorded in benefit data. In some cases, residential address may remain unchanged but appear different due to differences in spelling or errors.

For link 4c, an indicator of the number of residential address changes for the birth mother¹⁶ recorded in Health data is used where this is greater than the benefit derived count. This count is based on the maximum of the count of changes in meshblock recorded in PHO data and the count of changes in domicile code recorded in hospitalisation data over the previous year. Note there will be cases where changes in residential address are not able to be captured because the address is not able to be linked to a meshblock /domicile code or address was not recorded.

Parents or caregivers who transitioned between benefit and prison in the last 5 years

Short name: Benefit-prison transitions

Values: 1=yes, 2=no

This indicator is only considered for links 1c and 2c It is based on benefit data and set to one if at least one of the parents or caregivers transitioned between benefit and prison in the previous five years. Transitions are indicated by (i) an event code indicating release from prison as the reason for benefit grant or (ii) a reason code indicating entry to prison as the reason for cancellation of benefit.

Parents or caregivers with a Corrections history in the last 5 years

Short name: Corrections history

Values: 1-Non-custodial sentence, 2-custodial sentence for non-violent crimes, 3-custodial sentence for violent crimes, 4-no history

This categorical variable is based on Corrections data and indicates whether, in the last five years, the parents or caregivers had a sentence, whether at least one was custodial, and whether a least one custodial sentence was for violent crime defined as Australia New Zealand Standard Offence Classification (ANZSOC) offence groups 01 (homicide), 02 (assault), 03 (sexual assault), 06 (robbery, extortion), or 04,05 (other acts threatening persons) (Department of Corrections, 2012).

For all links, if children are identified through benefit data only, this variable relates to the caregivers (primary benefit recipient and their partner).

For links 3c and 4c, if children are identified through birth data only, this variable relates to the parents recorded in birth data.

For links 3c and 4c, if children are identified through birth data and benefit data, this variable relates to the parents recorded in Birth data, except in cases where the child is being cared for by caregiver/s who do not include the birth registration mother, in which case this indicator relates to the caregivers.

Neighborhood deprivation

This variable is a proxy for the deprivation level of the child's immediate neighborhood and is used in descriptive tables only. Using a geographical mapping application, addresses of children at birth from

¹⁶ Identified in Health data if she is one of the caregivers or parents listed on the birth registration (depending on whose data is being used).

benefit and birth registration data were linked to meshblocks. The meshblock is the smallest geographic unit for which statistical data is collected (approximately 50 households in each meshblock) and processed by Statistics New Zealand. An aggregated index of deprivation is calculated for each meshblock following every census (2006 was a census year). Called the New Zealand Deprivation index (NZDep), it reflects the following dimensions: income, home ownership, single parent family or not, employment, adult qualifications, living space, communication and transport. The scale of the NZDep index ranges from 1 to 10, with 10 being the most deprived (Salmond, Crampton, & Atkinson, 2007).

Community deprivation

Values: 1-low, 2-medium and 3-high

This variable is a proxy for the deprivation level of the child's community, where the boundaries of communities are defined by the Child Youth and Family (CYF) sites. There are 46 CYF sites operating in New Zealand. The variable is constructed for link 3c only.

The proportion of new born children in the study cohort within each CYF site who are from high deprivation neighbourhoods (NZDep 8, 9 or 10) is used to group CYF sites into low, medium and high deprivation communities with splits determined by tertiles of the distribution.

Address information is selected for each child using the rules set out for the neighborhood deprivation variable. Where residential address is missing or unable to be linked to a meshblock, community deprivation is set to medium.

Community vigilance

Values: 1-low, 2-medium and 3-high

This variable is intended to capture the level of vigilance in the community, proxied by the proportion of children who are notified to CYF. Boundaries of communities are defined by the Child Youth and Family (CYF) sites. The variable is constructed for link 3c only.

Within each CYF site the count of unique children in a cohort with a notification by age two is divided by the total number of children in that cohort and this ratio is used to group CYF sites into low, medium and high vigilance communities with splits determined by tertiles of the distribution.

Address information is selected for each child using the rules set out for the neighborhood deprivation variable. Where residential address is missing or unable to be linked to a meshblock, community vigilance is set to medium.

Note that if this variable were to be operationalized, it would need to be derived with a lag so that outcomes to age two could be observed.

Rate of investigation

Values: 1-low, 2-medium and 3-high

This variable is intended to capture the investigative responsiveness of CYF sites to notifications received, proxied by the proportion of children who are notified to CYF for whom there is an investigation or Child and Family Assessment. Boundaries of communities are defined by the Child Youth and Family (CYF) sites. The variable is constructed for link 3c only.

Within each CYF site the count of unique children in a representative cohort with an investigation or Child and Family Assessment is divided by count of unique children with a notification by age two and this ratio is used to group CYF sites into low, medium and high rate of investigation communities with splits determined by tertiles of the distribution.

Address information is selected for each child using the rules set out for the neighborhood deprivation variable. Where residential address is missing or unable to be linked to a meshblock, community vigilance is set to medium.

Note that if this variable were to be operationalized, it would need to be derived with a lag so that outcomes

to age two could be observed.

Community vigilance based on multiple counts

Values: 1-low, 2-medium and 3-high

This variable is intended to capture the level of vigilance in the community, proxied by the proportion of children who are notified to CYF. Boundaries of communities are defined by the Child Youth and Family (CYF) sites. The variable is constructed for link 3c only.

Within each CYF site the sum of multiple notifications by age two for children of representative cohort is divided by the total number of children in that cohort and this ratio is used to group CYF sites into low, medium and high vigilance communities with splits determined by tertiles of the distribution.

Address information is selected for each child using the rules set out for the neighborhood deprivation variable. Where residential address is missing or unable to be linked to a meshblock, community vigilance is set to medium.

Note that if this variable were to be operationalized, it would need to be derived with a lag so that outcomes to age two could be observed.

Rate of investigation based on multiple counts

Values: 1-low, 2-medium and 3-high

This variable is intended to capture the investigative responsiveness of CYF sites to notifications received, proxied by the proportion of children who are notified to CYF for whom there is an investigation or Child and Family Assessment. Boundaries of communities are defined by the Child Youth and Family (CYF) sites. The variable is constructed for link 3c only.

Within each CYF site on a representative cohort, the sum of multiple investigations or Child and Family Assessments are divided by sum of multiple notifications by age two and this ratio is used to group CYF sites into low, medium and high rate of investigation communities with splits determined by tertiles of the distribution.

Address information is selected for each child using the rules set out for the neighborhood deprivation variable. Where residential address is missing or unable to be linked to a meshblock, community vigilance is set to medium.

Note that if this variable were to be operationalized, it would need to be derived with a lag so that outcomes to age two could be observed.

Mother is smoker

Values: 1-yes, 2-no / unknown

This indicator is considered only for link 4c. It is set to 1 if hospitalisation data indicates the mother was a smoker at the child's birth.

Previous child not immunised and no decline

Short name: Other children immunisation history

Values: 1-yes, 2-no / unknown

This indicator is considered only for link 4c. It is based on National Immunisation Register data for other

children of the parents or caregivers identified through benefit and birth records, where the birth mother in health data is the same as one of those parents or caregivers identified through benefit and birth records. Coded “yes” where a previous child has immunisation status (excluding pneumococcal conjugate vaccine) “not complete” at ages 6, 8 and 12 months and there is no record of a parent or guardian declining immunisation.

Appendix 2 – Summary of model diagnostics

(a) Various algorithms tested

Model		Threshold (%)	Numbers to treat	AUR	AUR 95% CI Lower Limits	AUR 95% CI Upper limits	Misclassification rate (%)	Cumulative Lift	PPV by age 2 (%)	Sensitivity by age 2 (%)
Scored cohort 2007	Gradient boosting	3	1,868	85%	85%	86%	5%	5.7	13	17
	Dmine regression	3	1,868	87%	86%	88%	4%	8.7	20	26
	Neural network with 3 hidden units	3	1,868	87%	87%	88%	4%	8.1	19	24
	Neural network with 3 hidden units only with selected variables	3	1,868	86%	85%	87%	4%	7.9	18	24
	Neural network with 4 hidden units only with selected variables	3	1,868	86%	85%	87%	4%	7.9	18	24
	Partial Least squares	3	1,868	88%	87%	89%	4%	8.9	21	27
	Full regression	3	1,868	87%	86%	88%	4%	8.3	19	25
	Stepwise regression	3	1,868	87%	86%	88%	4%	8.3	19	25
	Backward regression	3	1,868	87%	86%	88%	4%	8.3	19	25
	Decision tree with max 2 branches and depth 6	3	1,868	87%	86%	88%	4%	8.1	19	24
	Decision tree with max 6 branches and depth 6	3	1,868	86%	84%	86%	4%	8.5	20	25
	Multilevel model	3	1,868	88%	87%	89%	4%	9.3	21	28
Training on cohort 2010	Gradient boosting			88%	87%	89%				
	DMINE regression			89%	88%	90%				
	Neural network with 3 hidden units			89%	89%	91%				
	Neural network with 3 hidden units only with selected variables			89%	88%	90%				
	Neural network with 4 hidden units only with selected variables			89%	88%	90%				
	Partial Least squares regression			91%	90%	91%				
	Full regression			90%	89%	91%				
	Stepwise regression			90%	89%	91%				
	Backward regression			90%	89%	91%				
	Decision tree with max 2 branches and depth 6			89%	88%	90%				
	Decision tree with max 6 branches and depth 6			88%	87%	89%				
	Multilevel model			90%	89%	91%				
Validation on cohort 2010	Gradient boosting			87%	86%	89%				
	DMINE regression			89%	88%	91%				
	Neural network with 3 hidden units			89%	88%	90%				
	Neural network with 3 hidden units only with selected variables			88%	87%	90%				
	Neural network with 4 hidden units only with selected variables			89%	87%	90%				
	Partial Least squares regression			90%	89%	91%				
	Full regression			89%	88%	91%				
	Stepwise regression			89%	88%	91%				
	Backward regression			89%	88%	91%				
	Decision tree with max 2 branches and depth 6			89%	87%	90%				
	Decision tree with max 6 branches and depth 6			88%	86%	89%				
	Multilevel model			90%	88%	91%				

(b) Base models on different cohorts and various administrative data and linkages

Model		Threshold (%)	Numbers to treat	AUR	AUR 95% CI Lower Limits	AUR 95% CI Upper limits	Misclassification rate (%)	Cumulative Lift	PPV by age 2 (%)	Sensitivity by age 2 (%)
Base model on different cohorts										
Scored on cohort 2007	Base model 2007	3	1,868	87%	86%	88%	4%	8.2	19	25
	Base model 2008	3	1,868	88%	87%	88%	4%	9.1	21	27
	Base model 2009	3	1,868	88%	87%	88%	4%	8.3	19	25
	Base model 2010	3	1,868	87%	86%	88%	4%	8.3	19	25
Training on cohort 2010	Base model 2007			88%	87%	89%				
	Base model 2008			89%	88%	90%				
	Base model 2009			90%	89%	91%				
	Base model 2010			90%	89%	91%				
Validation on cohort 2010	Base model 2007			87%	85%	89%				
	Base model 2008			87%	86%	89%				
	Base model 2009			89%	87%	90%				
	Base model 2010			89%	88%	91%				
Models to test for sensitivity to various administrative data and linkages										
Scored on cohort 2007	Model for cohort 2010 using link 1c	8	1,147	72%	70%	73%	12%	2.8	21	23
	Model for cohort 2010 using link 2c	8	1,145	72%	70%	74%	12%	2.6	20	21
	Model for cohort 2010 using link 4c	3	1,869	87%	86%	88%	4%	8.7	20	26
	Model for cohort 2010 using link 3	4	2,451	88%	88%	89%	5%	8.4	24	33
Training on cohort 2010	Model for cohort 2010 using link 1c			77%	76%	79%				
	Model for cohort 2010 using link 2c			78%	77%	79%				
	Model for cohort 2010 using link 4c			89%	88%	90%				
	Model for cohort 2010 using link 3			90%	89%	91%				
Validation on cohort 2010	Model for cohort 2010 using link 1c			74%	72%	77%				
	Model for cohort 2010 using link 2c			74%	71%	76%				
	Model for cohort 2010 using link 4c			88%	87%	90%				
	Model for cohort 2010 using link 3			89%	88%	90%				

(c) Various models to correct for disproportionality

Model		Threshold (%)	Top n	AUR	AUR 95%	AUR 95%	Misclassification rate (%)	Cumulative Lift	PPV by age 2 (%)	Sensitivity by age 2 (%)
					CI Lower Limits	CI Upper limits				
Scored on cohort 2007	Maori model	6	1,037	82%	81%	83%	8%	4.7	23	28
	“Non-Maori” ethnicity group not recorded	2	900	86%	85%	88%	3%	11.2	15	22
	Model where no other children with history of contact with CYF	2	1,186	86%	85%	87%	3%	8.4	13	17
	Model where there are other children with history of contact with CYF	17	503	62%	59%	64%	25%	1.6	27	27
	Model for children of benefit caregivers based on link 3c	8	1,155	73%	72%	75%	12%	2.9	23	23
	Model for children of non-benefit caregivers based on link	1	478	81%	78%	83%	1%	14.6	9	15
	Model for children of benefit caregivers based on link 4c	8	1,156	74%	73%	76%	12%	3.1	24	25
Model for children of benefit caregivers based on link 4c	1	478	78%	75%	81%	1%	11.5	7	11	
Training on cohort 2010	Maori model			84%	83%	86%				
	“Non-Maori” ethnicity group not recorded			91%	90%	93%				
	Model where no other children with history of contact with CYF			89%	88%	91%				
	Model where there are other children with history of contact with CYF			63%	59%	64%				
	Model for children of benefit caregivers based on link 3c			77%	75%	78%				
	Model for children of non-benefit caregivers based on link			83%	82%	86%				
	Model for children of benefit caregivers based on link 4c			78%	77%	79%				
Model for children of benefit caregivers based on link 4c			81%	80%	84%					
Validation on cohort 2010	Maori model			82%	80%	84%				
	“Non-Maori” ethnicity group not recorded			90%	88%	93%				
	Model where no other children with history of contact with CYF			88%	86%	90%				
	Model where there are other children with history of contact with CYF			64%	59%	64%				
	Model for children of benefit caregivers based on link 3c			76%	73%	78%				
	Model for children of non-benefit caregivers based on link			83%	81%	87%				
	Model for children of benefit caregivers based on link 4c			76%	74%	79%				
Model for children of benefit caregivers based on link 4c			81%	79%	85%					

(d) Models targeting different outcomes

Model		Threshold (%)	Top n	AUR	AUR 95% CI Lower Limits	AUR 95% CI Upper limits	Misclassification rate (%)	Cumulative Lift	PPV by age 2 (%)	Sensitivity by age 2 (%)
Scored on cohort 2007	Model to predict investigations or CFA	5	3,114	86%	86%	87%	7%	6.5	29	32
	Model to predict substantiated physical or sexual abuse or neglect (excluding emotional abuse findings)	1	623	87%	86%	88%	2%	12.1	12	12
	Model to predict substantiated physical abuse or hospitalisation for maltreatment or marker injury hospitalisation	1	623	75%	71%	78%	1%	7.9	2	8
	Model to predict notifications (including Police FV notifications and contact records)	8	4,982	86%	85%	86%	9%	5.1	38	41
Training on cohort 2010	Model to predict investigations or Child and Family Assessments			88%	87%	89%				
	Model to predict substantiated physical or sexual abuse or neglect (excluding emotional abuse findings)			90%	89%	91%				
	Model to predict substantiated physical abuse or hospitalisation for maltreatment or marker injury hospitalisation			77%	75%	80%				
	Model to predict notifications (including Police FV notifications and contact records)			87%	86%	87%				
Validation on cohort 2010	Model to predict investigations or Child and Family Assessments			87%	85%	89%				
	Model to predict substantiated physical or sexual abuse or neglect (excluding emotional abuse findings)			90%	88%	92%				
	Model to predict substantiated physical abuse or hospitalisation for maltreatment or marker injury hospitalisation			79%	76%	83%				
	Model to predict notifications (including Police FV notifications and contact records)			86%	85%	87%				
Scored on cohort 2007	Model with aggregated local level variables	3	1,868	88%	87%	89%	4%	8.9	21	27
	Model with no local variables	3	1,868	87%	86%	88%	4%	8.3	19	25
	Model with aggregated local level variables based on multiple counts of notifications and investigations	3	1,868	88%	87%	89%	4%	8.6	20	26
Training on cohort 2010	Model with aggregated local level variables			90%	89%	91%				
	Model with no local variables			90%	89%	91%				
	Model with aggregated local level variables based on multiple counts of notifications and investigations			90%	89%	91%				
Validation on cohort 2010	Model with aggregated local level variables			90%	88%	91%				
	Model with no local variables			89%	88%	91%				
	Model with aggregated local level variables based on multiple counts of notifications and investigations			90%	88%	91%				

(e) Multilevel models for different cohorts

	Model	Threshold (%)	Top n	AUR	AUR 95% CI Lower Limits	AUR 95% CI Upper limits	Misclassification rate (%)	Cumulative Lift	PPV by age 2 (%)	Sensitivity by age 2(%)
Scored on cohort 2007	Multilevel model for cohort 2008	3	1,868	88%	87%	89%	4%	9.0	21	27
	Multilevel model for cohort 2009	3	1,868	88%	87%	89%	4%	8.5	20	26
	Multilevel model for cohort 2010	3	1,868	88%	87%	89%	4%	9.3	21	28
Training on cohort 2010	Multilevel model for cohort 2008			90%	87%	89%				
	Multilevel model for cohort 2009			90%	89%	91%				
	Multilevel model for cohort 2010			90%	89%	91%				
Validation on cohort 2010	Multilevel model for cohort 2008			88%	87%	90%				
	Multilevel model for cohort 2009			89%	88%	90%				
	Multilevel model for cohort 2010			90%	88%	91%				

Appendix 3 – Summary of stepwise selection

Link	Model	Step	Variable	DF	Chi square (at entry)	Sig.	Validation error rate
3c	Base model for cohort 2010	1	Other children with care and protection history	1	2,766	<.0001	4,328
		2	Time on benefit in last 5 years	3	1,087	<.0001	3,813
		3	Caregiver with care and protection history	1	201	<.0001	3,698
		4	CYF service centre	42	210	<.0001	3,717
		5	Benefit caregiver is not birth registration parent	2	72	<.0001	3,705
		6	Mental health in last 5 years	4	88	<.0001	3,720
		7	Family violence	2	68	<.0001	3,710
		8	Caregiver's age	5	67	<.0001	3,683
		9	Correction history in the last 5 years	3	45	<.0001	3,652
		10	Single parent	2	36	<.0001	3,653
		11	Benefit address changes in last year	3	34	<.0001	3,657
		12	Behavioural or relationship difficulties as a child	1	21	<.0001	3,643
		13	Parenting demands	2	14	0.00	3,629
3c	Base model for cohort 2007	1	Other children with care and protection history	1	2,830	<.0001	3,724
		2	Time on benefit in last 5 years	3	892	<.0001	3,351
		3	Caregiver with care and protection history	1	204	<.0001	3,283
		4	Single parent	2	103	<.0001	3,275
		5	Correction history in the last 5 years	3	100	<.0001	3,274
3c	Base model for cohort 2008	1	Other children with care and protection history	1	3,377	<.0001	4,031
		2	Time on benefit in last 5 years	3	903	<.0001	3,622
		3	Caregiver with care and protection history	1	170	<.0001	3,603
		4	Mental health in last 5 years	4	78	<.0001	3,582
		5	Family violence	2	82	<.0001	3,560
		6	Single parent	2	57	<.0001	3,494
		7	Correction history in the last 5 years	3	53	<.0001	3,492
3c	Base model for cohort 2009	1	Other children with care and protection history	3	2391	<.0001	4,042
		2	Time on benefit in last 5 years	42	953	<.0001	3,648
		3	Caregiver with care and protection history	1	269	<.0001	3,620
		4	Mental health in last 5 years	4	131	<.0001	3,617
		5	CYF service centre	2	193	<.0001	3,623
		6	Parenting demands	3	68	<.0001	3,610
		7	Family violence	5	62	<.0001	3,595
		8	Correction history in the last 5 years	1	48	<.0001	3,593
		9	Benefit address changes in last year	2	52	<.0001	3,568
		10	Caregiver's age	3	56	<.0001	3,529
		11	Single parent	2	35	<.0001	3,479
		12	Benefit caregiver is not birth registration parent	2	15	0.00	3,469
1c	Model based on benefit and care protection data only	1	Other children with care and protection history	5	503	<.0001	2,775
		2	Caregiver with care and protection history	4	200	<.0001	2,721
		3	Benefit address changes in last year	3	65	<.0001	2,699
		4	Parenting demands	2	33	<.0001	2,697
		5	Time on benefit in last 5 years	3	43	<.0001	2,665
		6	Family violence	2	31	<.0001	2,665
		7	CYF service centre	42	87	<.0001	2,651
		8	Caregiver's age	5	18	0.00	2,648
		9	Mental health in last 5 years	4	24	<.0001	2,636
		10	Transition between benefit and prison	1	4	0.05	2,634
2c	Model based on benefit, care and protection data and Corrections data	1	Other children with care and protection history	5	490	<.0001	2,775
		2	Caregiver with care and protection history	4	215	<.0001	2,728
		3	Benefit address changes in last year	3	73	<.0001	2,713
		4	Time on benefit in last 5 years	3	48	<.0001	2,706
		5	Parenting demands	2	36	<.0001	2,687
		6	CYF service centre	42	111	<.0001	2,706
		7	Caregiver's age	5	24	0.00	2,711
		8	Mental health in last 5 years	4	24	<.0001	2,702
		9	Family violence	2	12	0.00	2,685
		10	Single parent	1	6	0.02	2,692
		11	Correction history in the last 5 years	3	10	0.02	2,686
		12	Behavioural or relationship difficulties as a child	1	1	0.23	2,684

Appendix 3 -continued

Link	Model	Step	Variable	DF	Chi square (at entry)	Sig.	Validation error rate
4c	Model based on benefit, care and protection, Corrections and Health data	1	Other children with care and protection history	1	308	<.0001	569
		2	Time on benefit in last 5 years	3	149	<.0001	507
		3	Caregiver with care and protection history	1	42	<.0001	502
		4	Mother is smoker	1	20	<.0001	500
		5	Caregiver's address changes in last year	3	12	0.01	500
		6	Mental health in last 5 years	4	11	0.03	493
3	Model for cohort 2010 using less conservative link 3	1	Other children with care and protection history	1	575	<.0001	903
		2	Time on benefit in last 5 years	3	263	<.0001	796
		3	Caregiver with care and protection history	1	43	<.0001	778
		4	Single parent	2	24	<.0001	779
		5	Family violence	2	13	0.00	777
		6	Correction history in the last 5 years	3	17	0.00	778
		7	Caregiver's age	5	20	0.00	783
		8	Benefit caregiver is not birth registration parent	2	9	0.01	776
		9	Mental health in last 5 years	4	10	0.04	769
3c	Model where no other children with history of contact with CYF	1	Time on benefit in last 5 years	3	1,355	<.0001	2,619
		2	Caregiver with care and protection history	1	240	<.0001	2,496
		3	Mental health in last 5 years	4	107	<.0001	2,498
		4	CYF service centre	42	167	<.0001	2,503
		5	Correction history in the last 5 years	3	83	<.0001	2,494
		6	Single parent	2	64	<.0001	2,437
		7	Caregiver's age	5	53	<.0001	2,391
		8	Family violence	2	30	<.0001	2,382
		9	Low birth weight or pre-term	1	18	<.0001	2,390
		10	Benefit address changes in last year	3	21	<.0001	2,394
		11	Benefit caregiver is not birth registration parent	2	23	<.0001	2,376
		12	Parenting demands	2	11	0.00	2,378
		13	Behavioural or relationship difficulties as a child	1	5	0.03	2,374
3c	Model where other children with history of contact with CYF	1	Family violence	2	48	<.0001	1,208
		2	Caregiver with care and protection history	1	38	<.0001	1,200
3c	Model for Maori ethnic group	1	Other children with care and protection history	1	683	<.0001	2,233
		2	Time on benefit in last 5 years	3	256	<.0001	2,088
		3	Caregiver with care and protection history	1	84	<.0001	2,034
		4	Benefit caregiver is not birth registration parent	2	54	<.0001	2,032
		5	CYF service centre	42	128	<.0001	2,060
		6	Parenting demands	2	32	<.0001	2,054
		7	Single parent	2	22	<.0001	2,032
		8	Mental health in last 5 years	4	26	<.0001	2,028
		9	Family violence	2	24	<.0001	2,015
		10	Caregiver's age	5	24	0.00	2,018
		11	Correction history in the last 5 years	3	20	0.00	2,017
		12	Benefit address changes in last year	3	8	0.04	2,013
3c	Model for "Non-Maori" and ethnic group not recorded	1	Family violence	2	1,807	<.0001	2,069
		2	Time on benefit in last 5 years	3	1,025	<.0001	1,743
		3	Caregiver with care and protection history	1	139	<.0001	1,677
		4	Mental health in last 5 years	4	115	<.0001	1,633
		5	CYF service centre	42	178	<.0001	1,611
		6	Caregiver's age	5	66	<.0001	1,616
		7	Other children with care and protection history	1	75	<.0001	1,588
		8	Single parent	2	25	<.0001	1,558
		9	Correction history in the last 5 years	3	28	<.0001	1,554
		10	Behavioural or relationship difficulties as a child	1	11	0.00	1,551

Appendix 3 -continued

Link	Model	Step	Variable	DF	Chi square (at entry)	Sig.	Validation error rate
3c	Model for beneficiaries	1	Other children with care and protection history	1	415	<.0001	2,814
		2	Caregiver with care and protection history	1	187	<.0001	2,749
		3	Time on benefit in last 5 years	3	50	<.0001	2,727
		4	Family violence	2	37	<.0001	2,717
		5	Correction history in the last 5 years	3	35	<.0001	2,707
		6	Caregiver's age	5	40	<.0001	2,706
		7	Mental health in last 5 years	4	40	<.0001	2,689
		8	Behavioural or relationship difficulties as a child	1	18	<.0001	2,691
		9	CYF service centre	42	80	0.00	2,667
		10	Benefit address changes in last year	3	20	0.00	2,670
		11	Benefit caregiver is not birth registration parent	2	22	<.0001	2,663
		12	Parenting demands	2	10	0.01	2,656
		13	Low birth weight or pre-term	1	3	0.07	2,655
3c	Model for non-Beneficiaries	1	Other children with care and protection history	1	2,032	<.0001	2,900
		2	Caregiver with care and protection history	1	703	<.0001	2,758
		3	Mental health in last 5 years	4	343	<.0001	2,729
		4	Time on benefit in last 5 years	3	249	<.0001	2,622
		5	Correction history in the last 5 years	3	155	<.0001	2,606
4c	Model for beneficiaries	1	Other children with care and protection history	1	424	<.0001	2,786
		2	Caregiver with care and protection history	1	172	<.0001	2,716
		3	Mental health in last 5 years	4	94	<.0001	2,674
		4	Family violence	2	53	<.0001	2,665
		5	Mother is smoker	1	43	<.0001	2,638
		6	Caregiver's age	5	42	<.0001	2,637
		7	Time on benefit in last 5 years	3	31	<.0001	2,619
		8	Correction history in the last 5 years	3	23	<.0001	2,619
		9	Benefit caregiver is not birth registration parent	2	24	<.0001	2,624
		10	CYF service centre	42	83	0.00	2,603
		11	Behavioural or relationship difficulties as a child	1	12	0.00	2,602
		12	Caregiver's address changes in last year	3	14	0.00	2,604
		13	Low birth weight or pre-term (MOH)	1	8	0.01	2,601
		14	Parenting demands	2	8	0.02	2,594
		15	Previous child not immunised and no decline	1	2	0.18	2,589
4c	Model for non-Beneficiaries	1	Other children with care and protection history	1	1,475	<.0001	2,950
		2	Time on benefit in last 5 years	3	602	<.0001	2,733
		3	Caregiver with care and protection history	1	197	<.0001	2,617
3c	Model to predict substantiated physical or sexual abuse or neglect (excluding emotional abuse findings)	1	Other children with care and protection history	1	1,451	<.0001	2,020
		2	Time on benefit in last 5 years	3	404	<.0001	1,837
		3	Caregiver with care and protection history	1	76	<.0001	1,786
		4	Benefit caregiver is not birth registration parent	2	55	<.0001	1,731
		5	Parenting demands	2	40	<.0001	1,723
3c	Model to predict investigations or Child and Family Assessments	1	Time on benefit in last 5 years	3	4,007	<.0001	6,623
		2	Other children with care and protection history	1	583	<.0001	6,360
		3	Caregiver with care and protection history	1	485	<.0001	6,198
		4	Single parent	2	163	<.0001	6,119
		5	Mental health in last 5 years	4	181	<.0001	6,079
		6	Caregiver's age	5	144	<.0001	6,055
3c	Model to predict substantiated physical or maltreatment or marker injury hospitalisation	1	Other children with care and protection history	1	486	<.0001	2,708
		2	Time on benefit in last 5 years	3	204	<.0001	2,596
		3	Caregiver with care and protection history	1	73	<.0001	2,560
		4	Benefit caregiver is not birth registration parent	2	34	<.0001	2,561
		5	Caregiver's age	5	32	<.0001	2,555
3c	Model to predict marker injury hospitalisation	1	Time on benefit in last 5 years	3	248	<.0001	1,820
		2	Behavioural or relationship difficulties as a child	1	54	<.0001	1,806
		3	Other children with care and protection history	1	23	<.0001	1,800

Appendix 3 -continued

Link	Model	Step	Variable	DF	Chi square (at entry)	Sig.	Validation error rate
3c	Model to predict notifications (including Police FV notifications and contact records)	1	Time on benefit in last 5 years	3	5,006	<.0001	8,618
		2	Caregiver with care and protection history	1	681	<.0001	8,332
		3	Other children with care and protection history	1	547	<.0001	8,068
		4	Single parent	2	213	<.0001	8,008
		5	Mental health in last 5 years	4	207	<.0001	7,950
		6	Correction history in the last 5 years	3	118	<.0001	7,936
		7	Caregiver's age	5	130	<.0001	7,893
		8	Family violence	2	65	<.0001	7,870
		9	Benefit address changes in last year	3	43	<.0001	7,852
		10	Parenting demands	2	31	<.0001	7,839
		11	Benefit caregiver is not birth registration parent	2	24	<.0001	7,819
		12	Behavioural or relationship difficulties as a child	1	10	0.00	7,814
		13	Low birth weight or pre-term	1	9	0.00	7,811
3c	Model with no local variables	1	Other children with care and protection history	1	2,766	<.0001	4,328
		2	Time on benefit in last 5 years	3	1,087	<.0001	3,813
		3	Caregiver with care and protection history	1	201	<.0001	3,698
		4	Benefit caregiver is not birth registration parent	2	75	<.0001	3,685
		5	Mental health in last 5 years	4	70	<.0001	3,688
		6	Family violence	2	65	<.0001	3,676
		7	Caregiver's age	5	64	<.0001	3,653
		8	Correction history in the last 5 years	3	52	<.0001	3,624
		9	Single parent	2	31	<.0001	3,623
		10	Benefit address changes in last year	3	29	<.0001	3,620
		11	Behavioural or relationship difficulties as a child	1	17	<.0001	3,608
		12	Parenting demands	2	13	0.00	3,597
		3c	Model with aggregated local level variables	1	Other children with care and protection history	1	2,766
2	Time on benefit in last 5 years			3	1,087	<.0001	3,813
3	Caregiver with care and protection history			1	201	<.0001	3,698
4	Benefit caregiver is not birth registration parent			2	75	<.0001	3,685
5	Mental health in last 5 years			4	70	<.0001	3,688
6	Family violence			2	65	<.0001	3,676
7	Caregiver's age			5	64	<.0001	3,653
8	Correction history in the last 5 years			3	52	<.0001	3,624
9	Single parent			2	31	<.0001	3,623
10	Benefit address changes in last year			3	29	<.0001	3,620
11	Behavioural or relationship difficulties as a child			1	17	<.0001	3,608
12	Parenting demands			2	13	0.00	3,597
13	Rate of Investigations			2	8	0.02	3,585
3c	Model with aggregated local level variables based on multiple counts of notifications and investigations	1	Other children with care and protection history	1	2,766	<.0001	4,328
		2	Time on benefit in last 5 years	3	1,087	<.0001	3,813
		3	Caregiver with care and protection history	1	201	<.0001	3,698
		4	Benefit caregiver is not birth registration parent	2	75	<.0001	3,685
		5	Mental health in last 5 years	4	70	<.0001	3,688
		6	Family violence	2	65	<.0001	3,676
		7	Caregiver's age	5	64	<.0001	3,653
		8	Correction history in the last 5 years	3	52	<.0001	3,624
		9	Single parent	2	31	<.0001	3,623
		10	Benefit address changes in last year	3	29	<.0001	3,620
		11	Behavioural or relationship difficulties as a child	1	17	<.0001	3,608
		12	Rate of Investigations	2	16	0.00	3,602
		13	Parenting demands	2	13	0.00	3,592

Appendix 4 – Summary of model estimates

(a) Base model for cohort 2010

Variable	Categories	Odds Ratio	Estimate	SE	Chi-Square	Sign.	95% CI Lower	95% CI Upper
Benefit address changes in last year	No address changes vs Missing/no benefit in last year	0.81	-0.21	0.07	8.1	0.00	-0.36	-0.07
Benefit address changes in last year	1 or 2 address changes vs Missing/no benefit in last year	0.74	-0.30	0.07	16.22	<.0001	-0.44	-0.15
Benefit address changes in last year	3 plus address changes vs Missing/no benefit in last year	1.43	0.36	0.10	12.92	0.00	0.16	0.55
Benefit caregiver is not birth registration parent	Yes vs No	2.02	0.70	0.14	26.13	<.0001	0.43	0.97
Benefit caregiver is not birth registration parent	No birth registration vs No	1.00	0.00	0.09	0	1.00	-0.17	0.17
Caregiver with care and protection history	Yes vs No	1.30	0.26	0.04	35.67	<.0001	0.17	0.34
Correction history in the last 5 years	Non-custodial sentence vs No history	1.06	0.06	0.08	0.45	0.50	-0.11	0.22
Correction history in the last 5 years	Custodial Sentence for non-violent crimes vs No history	0.99	-0.01	0.12	0.01	0.90	-0.25	0.22
Correction history in the last 5 years	Custodial Sentence for violent crimes vs No history	1.37	0.32	0.11	8.83	0.00	0.11	0.53
Caregiver's age	Under 20 vs 30_34/missing	1.48	0.39	0.09	20.25	<.0001	0.22	0.56
Caregiver's age	20-24 vs 30-34	0.85	-0.16	0.07	5.68	0.02	-0.29	-0.03
Caregiver's age	25-29 vs 30-34	0.87	-0.13	0.07	3.5	0.06	-0.28	0.01
Caregiver's age	35-39 vs 30-34	0.99	-0.01	0.11	0	0.95	-0.21	0.20
Caregiver's age	40 plus vs 30-34	1.20	0.18	0.14	1.55	0.21	-0.10	0.46
Behavioural or relationship difficulties as a child	Yes vs No	1.25	0.22	0.05	21.28	<.0001	0.13	0.32
Time on benefit in last 5 years	More than 80% vs no time	2.06	0.72	0.07	107.96	<.0001	0.59	0.86
Time on benefit in last 5 years	20<-80% vs no time	1.83	0.60	0.06	103.9	<.0001	0.49	0.72
Time on benefit in last 5 years	Up to 20% vs no time	1.09	0.09	0.08	1.18	0.28	-0.07	0.24
Mental health in last 5 years	Substance abuse vs No known issues	1.26	0.23	0.12	3.41	0.06	-0.01	0.47
Mental health in last 5 years	Persistent substance abuse issues vs No known issues	1.75	0.56	0.16	12.06	0.00	0.24	0.87
Mental health in last 5 years	Mental health issues other than substance abuse vs No known issues	0.72	-0.33	0.10	12.05	0.00	-0.52	-0.14
Mental health in last 5 years	Persistent mental health issues other than substance abuse vs No known issues	1.13	0.12	0.11	1.12	0.29	-0.10	0.34
Other children with care and protection history	Yes vs No	1.76	0.56	0.04	186.18	<.0001	0.48	0.64
Single parent	Yes vs Not single or unknown	1.17	0.15	0.05	9.5	0.00	0.06	0.25
Single parent	Single and no father listed vs not single or unknown	1.19	0.18	0.07	6.24	0.01	0.04	0.32
Parenting demands	High parenting demands vs no high parenting demands	1.14	0.13	0.05	7.2	0.01	0.04	0.23
Parenting demands	No other children vs no parenting demands	1.05	0.05	0.05	1.02	0.31	-0.05	0.16
Family violence	Events in one of the last 12 months vs no events	1.07	0.07	0.09	0.58	0.45	-0.11	0.25
Family violence	Events in more than one of the last 12 months vs no events	1.68	0.52	0.11	23.54	<.0001	0.31	0.73

(a) **Base model for cohort 2010-continued**

Variable	Categories	Odds Ratio	Estimate	SE	Chi-Square	Sign.	95% CI Lower	95% CI Upper
CYF Service centre	Blenheim vs Whangarei	1.24	0.21	0.43	0.25	0.62	-0.63	1.06
CYF Service centre	CYF Missing vs Whangarei	1.02	0.02	0.15	0.02	0.89	-0.27	0.31
CYF Service centre	Christchurch City vs Whangarei	0.60	-0.51	0.23	4.9	0.03	-0.96	-0.06
CYF Service centre	Clendon vs Whangarei	1.14	0.13	0.23	0.33	0.56	-0.32	0.58
CYF Service centre	Far North vs Whangarei	0.80	-0.22	0.22	1.03	0.31	-0.64	0.20
CYF Service centre	Gisborne vs Whangarei	1.30	0.26	0.20	1.75	0.19	-0.13	0.64
CYF Service centre	Grey Lynn vs Whangarei	1.50	0.41	0.18	4.95	0.03	0.05	0.77
CYF Service centre	Hastings vs Whangarei	0.76	-0.28	0.22	1.6	0.21	-0.72	0.15
CYF Service centre	Hauraki vs Whangarei	0.89	-0.11	0.27	0.18	0.67	-0.64	0.42
CYF Service centre	Hutt vs Whangarei	1.44	0.37	0.15	5.64	0.02	0.06	0.67
CYF Service centre	Manawatu vs Whangarei	0.69	-0.37	0.17	4.76	0.03	-0.69	-0.04
CYF Service centre	Mangere vs Whangarei	0.72	-0.32	0.19	3.02	0.08	-0.69	0.04
CYF Service centre	Manurewa vs Whangarei	0.98	-0.02	0.17	0.01	0.92	-0.35	0.31
CYF Service centre	Napier vs Whangarei	0.43	-0.83	0.32	6.6	0.01	-1.47	-0.20
CYF Service centre	Nelson vs Whangarei	2.08	0.73	0.23	9.77	0.00	0.27	1.19
CYF Service centre	Onehunga vs Whangarei	0.83	-0.18	0.26	0.52	0.47	-0.69	0.32
CYF Service centre	Orewa vs Whangarei	1.65	0.50	0.26	3.67	0.06	-0.01	1.01
CYF Service centre	Otago Urban vs Whangarei	1.29	0.25	0.26	0.91	0.34	-0.27	0.77
CYF Service centre	Otahuhu vs Whangarei	1.55	0.44	0.20	4.58	0.03	0.04	0.84
CYF Service centre	Otara vs Whangarei	0.96	-0.04	0.20	0.04	0.85	-0.43	0.36
CYF Service centre	Panmure vs Whangarei	1.27	0.24	0.24	1	0.32	-0.23	0.72
CYF Service centre	Papakura vs Whangarei	0.69	-0.37	0.21	3.15	0.08	-0.77	0.04
CYF Service centre	Papanui vs Whangarei	0.93	-0.08	0.20	0.15	0.70	-0.47	0.32
CYF Service centre	Porirua vs Whangarei	0.42	-0.87	0.29	9.03	0.00	-1.44	-0.30
CYF Service centre	Pukekohe vs Whangarei	1.13	0.12	0.24	0.27	0.60	-0.34	0.59
CYF Service centre	Rotorua vs Whangarei	1.42	0.35	0.17	4.38	0.04	0.02	0.69
CYF Service centre	South Canterbury vs Whangarei	0.60	-0.52	0.29	3.11	0.08	-1.09	0.06
CYF Service centre	Southern Rural vs Whangarei	0.89	-0.12	0.32	0.13	0.72	-0.74	0.51
CYF Service centre	Southland vs Whangarei	0.66	-0.41	0.25	2.78	0.10	-0.89	0.07
CYF Service centre	Sydenham vs Whangarei	0.99	-0.01	0.23	0	0.95	-0.47	0.44
CYF Service centre	Takapuna vs Whangarei	1.55	0.44	0.20	4.96	0.03	0.05	0.82
CYF Service centre	Taranaki vs Whangarei	0.73	-0.32	0.23	1.83	0.18	-0.78	0.14
CYF Service centre	Tauranga vs Whangarei	1.64	0.49	0.16	9.44	0.00	0.18	0.81
CYF Service centre	Waikato East vs Whangarei	0.62	-0.48	0.18	6.65	0.01	-0.84	-0.11
CYF Service centre	Waikato West vs Whangarei	0.98	-0.02	0.15	0.02	0.89	-0.31	0.27
CYF Service centre	Wairarapa vs Whangarei	0.89	-0.12	0.36	0.11	0.74	-0.82	0.59
CYF Service centre	Waitakere vs Whangarei	0.59	-0.53	0.19	7.6	0.01	-0.90	-0.15
CYF Service centre	Wellington vs Whangarei	0.60	-0.51	0.35	2.03	0.15	-1.20	0.19
CYF Service centre	West Coast vs Whangarei	0.93	-0.07	0.44	0.03	0.87	-0.94	0.79
CYF Service centre	Westgate vs Whangarei	1.46	0.38	0.16	5.28	0.02	0.06	0.70
CYF Service centre	Whakatane vs Whangarei	5.59	1.72	0.17	106.9	<.0001	1.40	2.05
CYF Service centre	Whanganui vs Whangarei	0.95	-0.05	0.21	0.05	0.82	-0.47	0.37

(b) Base model for cohort 2007

Variable	Categories	Odds Ratio	Estimate	SE	Chi-Square	Sign.	95% CI Lower	95% CI Upper
Caregiver with care and protection history	Yes vs No	1.56	0.44	0.04	145.34	<.0001	-1.16	0.86
Correction history in the last 5 years	Non-custodial sentence vs No history	1.41	0.34	0.09	14.84	0.00	-0.04	0.44
Correction history in the last 5 years	Custodial Sentence for non-violent crimes vs No history	0.95	-0.05	0.13	0.13	0.72	-0.47	0.35
Correction history in the last 5 years	Custodial Sentence for violent crimes vs No history	1.28	0.24	0.11	4.6	0.03	0.09	1.04
Time on benefit in last 5 years	More than 80% vs no time	2.04	0.71	0.07	111.54	<.0001	-1.17	0.01
Time on benefit in last 5 years	20<-80% vs no time	1.72	0.54	0.06	77.9	<.0001	0.28	1.17
Time on benefit in last 5 years	Up to 20% vs no time	0.99	-0.01	0.08	0.03	0.86	-0.45	0.40
Other children with care and protection history	Yes vs No	2.13	0.76	0.04	332.13	<.0001	-1.06	-0.07
Single parent	Yes vs Not single or unknown	1.10	0.10	0.05	3.77	0.05	0.97	2.11
Single parent	Single and no father listed vs not single or unknown	1.71	0.54	0.07	66.21	<.0001	0.23	0.87

(c) Base model for cohort 2008

Variable	Categories	Odds Ratio	Estimate	SE	Chi-Square	Sign.	95% CI Lower	95% CI Upper
Caregiver with care and protection history	Yes vs No	1.49	0.40	0.04	129.66	<.0001	-1.73	0.55
Correction history in the last 5 years	Non-custodial sentence vs No history	1.40	0.34	0.08	15.98	<.0001	-0.07	0.39
Correction history in the last 5 years	Custodial Sentence for non-violent crimes vs No history	1.30	0.26	0.12	4.47	0.03	-0.28	0.59
Correction history in the last 5 years	Custodial Sentence for violent crimes vs No history	0.74	-0.30	0.12	6.72	0.01	-0.01	0.86
Time on benefit in last 5 years	More than 80% vs no time	2.07	0.73	0.07	118.16	<.0001	0.11	0.96
Time on benefit in last 5 years	20<-80% vs no time	1.63	0.49	0.06	69.04	<.0001	-0.86	0.14
Time on benefit in last 5 years	Up to 20% vs no time	1.13	0.12	0.08	2.61	0.11	0.09	0.81
Mental health in last 5 years	Substance abuse vs No known issues	1.15	0.14	0.12	1.33	0.25	-0.43	0.48
Mental health in last 5 years	Persistent substance abuse issues vs No known issues	1.67	0.52	0.17	8.79	0.00	-0.37	0.75
Mental health in last 5 years	Mental health issues other than substance abuse vs No known issues	0.82	-0.20	0.10	3.97	0.05	0.16	0.81
Mental health in last 5 years	Persistent mental health issues other than substance abuse vs No known issues	1.17	0.16	0.13	1.6	0.21	-0.47	0.21
Other children with care and protection history	Yes vs No	1.99	0.69	0.04	275.55	<.0001	-0.34	0.43
Single parent	Yes vs Not single or unknown	1.24	0.22	0.05	19.63	<.0001	-0.54	0.28
Single parent	Single and no father listed vs not single or unknown	1.20	0.19	0.07	7.55	0.01	-1.30	-0.17
Family violence	Events in one of the last 12 months vs no events	1.21	0.19	0.12	2.48	0.12	0.00	1.12
Family violence	Events in more than one of the last 12 months vs no events	1.89	0.64	0.16	16.33	<.0001	-1.65	-0.33

(d) Base model for cohort 2009

Variable	Categories	Odds Ratio	Estimate	SE	Chi-Square	Sign.	95% CI Lower	95% CI Upper
Benefit address changes in last year	No address changes vs Missing/no benefit in last year	0.68	-0.39	0.07	27.1	<.0001	-0.54	-0.25
Benefit address changes in last year	1 or 2 address changes vs Missing/no benefit in last year	1.01	0.01	0.07	0.0	0.94	-0.14	0.14
Benefit address changes in last year	3 plus address changes vs Missing/no benefit in last year	1.59	0.47	0.10	23.6	<.0001	0.28	0.66
Benefit caregiver is not birth registration parent	Yes vs No	1.52	0.42	0.14	8.9	0.00	0.15	0.70
Benefit caregiver is not birth registration parent	No birth registration vs No	0.92	-0.08	0.09	0.8	0.38	-0.24	0.12
Caregiver with care and protection history	Yes vs No	1.42	0.35	0.04	77.5	<.0001	0.21	0.39
Correction history in the last 5 years	Non-custodial sentence vs No history	1.11	0.11	0.09	1.6	0.21	-0.06	0.28
Correction history in the last 5 years	Custodial Sentence for non-violent crimes vs No history	1.21	0.19	0.12	2.7	0.10	-0.06	0.40
Correction history in the last 5 years	Custodial Sentence for violent crimes vs No history	1.18	0.16	0.10	2.8	0.10	-0.02	0.37
Caregiver's age	Under 20 vs 30_34/missing	1.75	0.56	0.09	40.6	<.0001	0.37	0.72
Caregiver's age	20-24 vs 30-34	0.89	-0.12	0.07	2.9	0.09	-0.27	0.01
Caregiver's age	25-29 vs 30-34	0.97	-0.04	0.07	0.2	0.63	-0.18	0.10
Caregiver's age	35-39 vs 30-34	0.82	-0.20	0.10	3.6	0.06	-0.40	0.00
Caregiver's age	40 plus vs 30-34	0.88	-0.13	0.15	0.7	0.39	-0.40	0.19
Time on benefit in last 5 years	More than 80% vs no time	2.10	0.74	0.07	113.5	<.0001	0.04	0.23
Time on benefit in last 5 years	20<-80% vs no time	1.86	0.62	0.06	112.7	<.0001	0.60	0.87
Time on benefit in last 5 years	Up to 20% vs no time	0.85	-0.16	0.08	3.9	0.05	0.48	0.71
Mental health in last 5 years	Substance abuse vs No known issues	1.13	0.12	0.11	1.1	0.29	-0.29	0.02
Mental health in last 5 years	Persistent substance abuse issues vs No known issues	2.08	0.73	0.15	22.7	<.0001	-0.11	0.34
Mental health in last 5 years	Mental health issues other than substance abuse vs No known issues	0.90	-0.11	0.09	1.3	0.25	0.43	1.03
Mental health in last 5 years	Persistent mental health issues other than substance abuse vs No known issues	0.99	-0.01	0.12	0.0	0.96	-0.29	0.07
Other children with care and protection history	Yes vs No	1.76	0.57	0.04	167.3	<.0001	-0.25	0.23
Single parent	Yes vs Not single or unknown	1.23	0.21	0.05	16.4	<.0001	0.48	0.65
Single parent	Single and no father listed vs not single or unknown	1.08	0.08	0.07	1.2	0.28	0.10	0.30
Parenting demands	High parenting demands vs no high parenting demands	1.17	0.15	0.05	9.5	0.00	-0.04	0.26
Parenting demands	No other children vs no parenting demands	1.28	0.25	0.05	22.2	<.0001	0.08	0.30
Family violence	Events in one of the last 12 months vs no events	0.67	-0.40	0.11	14.5	0.00	0.04	0.24
Family violence	Events in more than one of the last 12 months vs no events	2.52	0.92	0.12	59.5	<.0001	0.14	0.35
CYF service centre	Parameter estimates available upon request							

(e) Model based on benefit and care protection data only

Variable	Categories	Odds Ratio	Estimate	SE	Chi-Square	Sign.	95% CI Lower	95% CI Upper
Benefit address changes in last year	No address changes vs Missing/no benefit in last year	1.01	0.01	0.10	0.0	0.91	-0.18	0.20
Benefit address changes in last year	1 or 2 address changes vs Missing/no benefit in last year	1.28	0.25	0.09	6.9	0.01	0.06	0.43
Benefit address changes in last year	3 plus address changes vs Missing/no benefit in last year	1.87	0.63	0.10	36.7	<.0001	0.43	0.84
Caregiver with care and protection history	Police FV notification or contact vs No history	0.84	-0.18	0.17	1.1	0.30	-0.51	0.17
Caregiver with care and protection history	Notifications and investigations only vs No history	1.02	0.02	0.09	0.1	0.81	-0.16	0.20
Caregiver with care and protection history	Substantiated findings of maltreatment vs No history	1.17	0.16	0.10	2.7	0.10	-0.03	0.35
Caregiver with care and protection history	Placement or care entry vs No history	1.48	0.39	0.09	19.3	<.0001	0.23	0.57
Caregiver's age	Under 20 vs 30_34/missing	1.52	0.42	0.09	20.7	<.0001	0.22	0.58
Caregiver's age	20-24 vs 30-34	1.10	0.10	0.07	2.0	0.16	-0.04	0.23
Caregiver's age	25-29 vs 30-34	0.84	-0.17	0.08	4.8	0.03	-0.33	-0.02
Caregiver's age	35-39 vs 30-34	0.88	-0.12	0.12	1.1	0.30	-0.35	0.12
Caregiver's age	40 plus vs 30-34	0.94	-0.07	0.15	0.2	0.65	-0.36	0.22
Time on benefit in last 5 years	More than 80% vs no time	1.59	0.46	0.09	27.3	<.0001	0.29	0.63
Time on benefit in last 5 years	20<-80% vs no time	1.31	0.27	0.08	10.2	0.00	0.11	0.44
Time on benefit in last 5 years	Up to 20% vs no time	0.86	-0.15	0.10	2.0	0.16	-0.35	0.06
Mental health in last 5 years	Substance abuse vs No known issues	1.13	0.12	0.16	0.6	0.44	-0.15	0.45
Mental health in last 5 years	Persistent substance abuse issues vs No known issues	1.65	0.50	0.21	5.9	0.02	0.13	0.94
Mental health in last 5 years	Mental health issues other than substance abuse vs No known issues	0.73	-0.32	0.11	7.6	0.01	-0.56	-0.12
Mental health in last 5 years	Persistent mental health issues other than substance abuse vs No known issues	1.04	0.04	0.13	0.1	0.74	-0.24	0.29
Other children with care and protection history	Police FV notification or contact vs No history	0.89	-0.11	0.14	0.7	0.42	-0.39	0.16
Other children with care and protection history	Notifications only vs No history	1.26	0.23	0.11	4.3	0.04	0.01	0.45
Other children with care and protection history	Notifications and investigations only vs No history	1.10	0.09	0.12	0.6	0.44	-0.14	0.33
Other children with care and protection history	Substantiated findings of maltreatment vs No history	1.63	0.49	0.09	31.6	<.0001	0.32	0.66
Other children with care and protection history	Placement or care entry vs No history	1.22	0.20	0.14	2.2	0.14	-0.05	0.48
Transition between benefit and prison	Yes vs No	1.15	0.14	0.07	3.7	0.05	-0.05	0.16
Parenting demands	High parenting demands vs no high parenting demands	1.06	0.06	0.05	1.2	0.27	0.12	0.37
Parenting demands	No other children vs no parenting demands	1.27	0.24	0.06	14.6	0.00	-0.18	0.20
Family violence	Events in one of the last 12 months vs no events	1.01	0.01	0.10	0.0	0.90	0.18	0.60
Family violence	Events in more than one of the last 12 months vs no events	1.47	0.39	0.11	13.6	0.00	0.18	0.59

(e) Model based on benefit and care protection data only -Continued

Variable	Categories	Odds Ratio	Estimate	SE	Chi-Square	Sign.	95% CI Lower	95% CI Upper
CYF Service centre	Blenheim vs Whangarei	1.83	0.61	0.38	2.6	0.11	-0.16	1.33
CYF Service centre	CYF Missing vs Whangarei	0.82	-0.20	0.17	1.3	0.25	-0.54	0.14
CYF Service centre	Christchurch City vs Whangarei	0.95	-0.05	0.22	0.1	0.82	-0.47	0.37
CYF Service centre	Clendon vs Whangarei	1.08	0.08	0.23	0.1	0.74	-0.39	0.53
CYF Service centre	Far North vs Whangarei	1.03	0.03	0.22	0.0	0.90	-0.40	0.45
CYF Service centre	Gisborne vs Whangarei	1.42	0.35	0.22	2.6	0.10	-0.07	0.77
CYF Service centre	Grey Lynn vs Whangarei	1.46	0.38	0.21	3.4	0.07	-0.03	0.77
CYF Service centre	Hastings vs Whangarei	0.76	-0.27	0.23	1.4	0.24	-0.72	0.18
CYF Service centre	Hauraki vs Whangarei	1.20	0.19	0.30	0.4	0.53	-0.40	0.77
CYF Service centre	Hutt vs Whangarei	1.90	0.64	0.16	15.3	<.0001	0.31	0.95
CYF Service centre	Manawatu vs Whangarei	0.78	-0.25	0.18	2.0	0.16	-0.60	0.10
CYF Service centre	Mangere vs Whangarei	0.94	-0.06	0.20	0.1	0.76	-0.46	0.34
CYF Service centre	Manurewa vs Whangarei	1.00	0.00	0.20	0.0	0.99	-0.39	0.38
CYF Service centre	Napier vs Whangarei	0.44	-0.82	0.36	5.0	0.02	-1.54	-0.11
CYF Service centre	Nelson vs Whangarei	1.18	0.16	0.30	0.3	0.59	-0.44	0.75
CYF Service centre	Onehunga vs Whangarei	0.82	-0.20	0.37	0.3	0.59	-0.92	0.51
CYF Service centre	Orewa vs Whangarei	1.49	0.40	0.32	1.5	0.21	-0.24	1.02
CYF Service centre	Otago Urban vs Whangarei	0.82	-0.20	0.31	0.4	0.52	-0.81	0.40
CYF Service centre	Otahuhu vs Whangarei	1.21	0.19	0.22	0.7	0.39	-0.25	0.63
CYF Service centre	Otara vs Whangarei	0.98	-0.03	0.22	0.0	0.91	-0.46	0.40
CYF Service centre	Panmure vs Whangarei	1.21	0.19	0.24	0.6	0.43	-0.28	0.66
CYF Service centre	Papakura vs Whangarei	1.01	0.01	0.21	0.0	0.98	-0.41	0.42
CYF Service centre	Papanui vs Whangarei	0.99	-0.01	0.23	0.0	0.98	-0.44	0.45
CYF Service centre	Porirua vs Whangarei	0.75	-0.29	0.24	1.5	0.22	-0.75	0.18
CYF Service centre	Pukekohe vs Whangarei	1.76	0.56	0.27	4.5	0.03	0.05	1.09
CYF Service centre	Rotorua vs Whangarei	1.02	0.02	0.19	0.0	0.93	-0.35	0.39
CYF Service centre	South Canterbury vs Whangarei	1.05	0.05	0.29	0.0	0.86	-0.52	0.61
CYF Service centre	Southern Rural vs Whangarei	0.70	-0.36	0.42	0.7	0.39	-1.18	0.47
CYF Service centre	Southland vs Whangarei	0.72	-0.32	0.30	1.2	0.28	-0.87	0.29
CYF Service centre	Sydenham vs Whangarei	0.59	-0.53	0.27	3.8	0.05	-1.06	0.01
CYF Service centre	Takapuna vs Whangarei	1.21	0.19	0.25	0.6	0.46	-0.32	0.67
CYF Service centre	Taranaki vs Whangarei	0.46	-0.77	0.31	6.2	0.01	-1.37	-0.15
CYF Service centre	Tauranga vs Whangarei	1.23	0.21	0.18	1.4	0.24	-0.14	0.56
CYF Service centre	Waikato East vs Whangarei	0.97	-0.03	0.19	0.0	0.88	-0.38	0.34
CYF Service centre	Waikato West vs Whangarei	0.92	-0.08	0.16	0.3	0.60	-0.40	0.23
CYF Service centre	Wairarapa vs Whangarei	0.78	-0.25	0.35	0.5	0.48	-0.94	0.42
CYF Service centre	Waitakere vs Whangarei	1.21	0.19	0.20	1.0	0.32	-0.20	0.57
CYF Service centre	Wellington vs Whangarei	0.64	-0.45	0.39	1.3	0.25	-1.20	0.33
CYF Service centre	West Coast vs Whangarei	0.95	-0.05	0.53	0.0	0.92	-1.09	0.99
CYF Service centre	Westgate vs Whangarei	1.33	0.29	0.18	2.5	0.11	-0.07	0.64
CYF Service centre	Whakatane vs Whangarei	2.41	0.88	0.18	24.6	<.0001	0.53	1.23
CYF Service centre	Whanganui vs Whangarei	0.62	-0.47	0.26	3.3	0.07	-0.96	0.06

(f) Model based on benefit, care and protection data and Corrections data

Variable	Categories	Odds Ratio	Estimate	SE	Chi-Square	Sign.	95% CI Lower	95% CI Upper
Benefit address changes in last year	No address changes vs Missing/no benefit in last year	0.94	-0.06	0.10	0.4	0.55	-0.25	0.13
Benefit address changes in last year	1 or 2 address changes vs Missing/no benefit in last year	1.28	0.25	0.09	7.2	0.01	0.07	0.42
Benefit address changes in last year	3 plus address changes vs Missing/no benefit in last year	1.85	0.62	0.10	37.2	<.0001	0.42	0.82
Caregiver with care and protection history	Police FV notification or contact vs No history	0.93	-0.07	0.16	0.2	0.65	-0.41	0.21
Caregiver with care and protection history	Notifications and investigations only vs No history	1.13	0.12	0.09	1.9	0.17	-0.03	0.31
Caregiver with care and protection history	Substantiated findings of maltreatment vs No history	1.05	0.05	0.10	0.3	0.60	-0.13	0.24
Caregiver with care and protection history	Placement or care entry vs No history	1.39	0.33	0.09	12.5	0.00	0.19	0.53
Correction history in the last 5 years	Non-custodial sentence vs No history	1.18	0.16	0.12	1.8	0.18	-0.07	0.41
Correction history in the last 5 years	Custodial Sentence for non-violent crimes vs No history	1.07	0.06	0.18	0.1	0.71	-0.28	0.41
Correction history in the last 5 years	Custodial Sentence for violent crimes vs No history	1.00	0.00	0.14	0.0	0.99	-0.27	0.28
Caregiver's age	Under 20 vs 30_34/missing	1.59	0.46	0.10	23.7	<.0001	0.28	0.65
Caregiver's age	20-24 vs 30-34	1.17	0.15	0.07	4.6	0.03	0.02	0.30
Caregiver's age	25-29 vs 30-34	0.89	-0.12	0.08	2.1	0.15	-0.28	0.04
Caregiver's age	35-39 vs 30-34	0.94	-0.06	0.12	0.3	0.60	-0.30	0.18
Caregiver's age	40 plus vs 30-34	0.76	-0.28	0.16	3.1	0.08	-0.59	0.03
Behavioural or relationship difficulties as a child	Yes vs No	1.07	0.07	0.06	1.5	0.23	0.29	0.64
Time on benefit in last 5 years	More than 80% vs no time	1.59	0.46	0.09	26.6	<.0001	0.11	0.44
Time on benefit in last 5 years	20<-80% vs no time	1.32	0.28	0.08	11.0	0.00	-0.51	-0.07
Time on benefit in last 5 years	Up to 20% vs no time	0.75	-0.29	0.11	6.8	0.01	-0.03	0.57
Mental health in last 5 years	Substance abuse vs No known issues	1.30	0.26	0.15	3.0	0.08	-0.03	0.79
Mental health in last 5 years	Persistent substance abuse issues vs No known issues	1.45	0.37	0.21	3.2	0.08	-0.58	-0.13
Mental health in last 5 years	Mental health issues other than substance abuse vs No known issues	0.70	-0.36	0.11	9.6	0.00	-0.16	0.36
Mental health in last 5 years	Persistent mental health issues other than substance abuse vs No known issues	1.11	0.10	0.13	0.6	0.43	-0.29	0.25
Caregiver with care and protection history	Police FV notification or contact vs No history	0.98	-0.02	0.14	0.0	0.90	-0.04	0.41
Caregiver with care and protection history	Notifications only vs No history	1.20	0.18	0.11	2.6	0.11	-0.09	0.38
Caregiver with care and protection history	Notifications and investigations only vs No history	1.16	0.15	0.12	1.5	0.23	0.32	0.66
Caregiver with care and protection history	Substantiated findings of maltreatment vs No history	1.63	0.49	0.09	32.6	<.0001	-0.20	0.36
Caregiver with care and protection history	Placement or care entry vs No history	1.08	0.08	0.14	0.3	0.59	0.04	0.25
Single parent	Yes vs No	1.16	0.15	0.05	7.5	0.01	0.04	0.25
Parenting demands	High parenting demands vs no high parenting demands	1.16	0.15	0.05	7.3	0.01	0.02	0.26
Parenting demands	No other children vs no parenting demands	1.15	0.14	0.06	4.8	0.03	-0.13	0.25
Family violence	Events in ne of the last 12 months vs no events	1.06	0.06	0.10	0.4	0.54	-0.02	0.42
Family violence	Events in more than one of the last 12 months vs no events	1.22	0.20	0.11	3.2	0.08	-0.02	0.42

(f) Model based on benefit, care and protection data and Corrections data-continued

Variable	Categories	Odds Ratio	Estimate	SE	Chi-Square	Sign.	95% CI Lower	95% CI Upper
CYF Service centre	Blenheim vs Whangarei	1.84	0.61	0.38	2.6	0.11	-0.13	1.36
CYF Service centre	CYF Missing vs Whangarei	1.00	0.00	0.19	0.0	0.99	-0.37	0.37
CYF Service centre	Christchurch City vs Whangarei	0.88	-0.13	0.23	0.3	0.56	-0.56	0.33
CYF Service centre	Clendon vs Whangarei	1.06	0.06	0.24	0.1	0.82	-0.41	0.53
CYF Service centre	Far North vs Whangarei	0.91	-0.09	0.22	0.2	0.68	-0.53	0.34
CYF Service centre	Gisborne vs Whangarei	1.34	0.30	0.22	1.8	0.18	-0.13	0.72
CYF Service centre	Grey Lynn vs Whangarei	1.30	0.26	0.22	1.4	0.23	-0.16	0.68
CYF Service centre	Hastings vs Whangarei	0.80	-0.22	0.22	1.0	0.32	-0.65	0.21
CYF Service centre	Hauraki vs Whangarei	1.26	0.23	0.30	0.6	0.43	-0.34	0.81
CYF Service centre	Hutt vs Whangarei	2.24	0.81	0.16	26.8	<.0001	0.50	1.11
CYF Service centre	Manawatu vs Whangarei	0.80	-0.23	0.18	1.6	0.21	-0.59	0.12
CYF Service centre	Mangere vs Whangarei	0.93	-0.07	0.21	0.1	0.72	-0.49	0.33
CYF Service centre	Manurewa vs Whangarei	1.56	0.45	0.17	7.2	0.01	0.12	0.77
CYF Service centre	Napier vs Whangarei	0.27	-1.32	0.45	8.5	0.00	-2.20	-0.43
CYF Service centre	Nelson vs Whangarei	1.69	0.52	0.27	3.8	0.05	0.00	1.06
CYF Service centre	Onehunga vs Whangarei	1.01	0.01	0.32	0.0	0.99	-0.63	0.64
CYF Service centre	Orewa vs Whangarei	0.87	-0.14	0.40	0.1	0.74	-0.92	0.65
CYF Service centre	Otago Urban vs Whangarei	0.83	-0.19	0.30	0.4	0.53	-0.77	0.40
CYF Service centre	Otahuhu vs Whangarei	0.79	-0.24	0.25	0.9	0.34	-0.73	0.25
CYF Service centre	Otara vs Whangarei	0.95	-0.06	0.22	0.1	0.80	-0.49	0.38
CYF Service centre	Panmure vs Whangarei	1.18	0.17	0.25	0.5	0.50	-0.32	0.64
CYF Service centre	Papakura vs Whangarei	1.02	0.02	0.21	0.0	0.94	-0.39	0.42
CYF Service centre	Papanui vs Whangarei	1.07	0.07	0.22	0.1	0.76	-0.36	0.51
CYF Service centre	Porirua vs Whangarei	0.62	-0.48	0.26	3.6	0.06	-0.98	0.02
CYF Service centre	Pukekohe vs Whangarei	1.70	0.53	0.27	4.0	0.05	0.01	1.05
CYF Service centre	Rotorua vs Whangarei	1.16	0.15	0.19	0.7	0.42	-0.22	0.51
CYF Service centre	South Canterbury vs Whangarei	1.24	0.21	0.28	0.6	0.45	-0.34	0.77
CYF Service centre	Southern Rural vs Whangarei	0.73	-0.32	0.42	0.6	0.45	-1.15	0.51
CYF Service centre	Southland vs Whangarei	0.78	-0.25	0.28	0.8	0.37	-0.80	0.29
CYF Service centre	Sydenham vs Whangarei	0.78	-0.25	0.26	0.9	0.34	-0.74	0.26
CYF Service centre	Takapuna vs Whangarei	1.03	0.03	0.28	0.0	0.92	-0.52	0.56
CYF Service centre	Taranaki vs Whangarei	0.47	-0.76	0.31	5.8	0.02	-1.37	-0.14
CYF Service centre	Tauranga vs Whangarei	0.98	-0.02	0.19	0.0	0.91	-0.39	0.34
CYF Service centre	Waikato East vs Whangarei	0.80	-0.22	0.20	1.2	0.27	-0.62	0.18
CYF Service centre	Waikato West vs Whangarei	0.96	-0.04	0.16	0.1	0.79	-0.36	0.27
CYF Service centre	Wairarapa vs Whangarei	0.61	-0.50	0.38	1.7	0.19	-1.26	0.24
CYF Service centre	Waitakere vs Whangarei	1.26	0.23	0.19	1.5	0.22	-0.14	0.61
CYF Service centre	Wellington vs Whangarei	0.72	-0.32	0.37	0.8	0.38	-1.06	0.39
CYF Service centre	West Coast vs Whangarei	1.28	0.24	0.50	0.2	0.63	-0.75	1.22
CYF Service centre	Westgate vs Whangarei	1.69	0.52	0.17	9.6	0.00	0.19	0.86
CYF Service centre	Whakatane vs Whangarei	2.30	0.83	0.18	21.3	<.0001	0.48	1.19
CYF Service centre	Whanganui vs Whangarei	0.67	-0.41	0.26	2.4	0.12	-0.91	0.10

(g) Model for cohort 2010 using less conservative link 3

Variable	Categories	Odds Ratio	Estimate	SE	Chi-Square	Sign.	95% CI Lower	95% CI Upper
Benefit caregiver is not birth registration parent	Yes vs No	0.88	-0.13	0.40	0.1	0.75	-0.93	0.64
Benefit caregiver is not birth registration parent	No birth registration vs No	1.41	0.35	0.22	2.5	0.12	-0.11	0.75
Caregiver with care and protection history	Yes vs No	1.34	0.29	0.08	13.7	0.00	0.17	0.47
Correction history in the last 5 years	Non-custodial sentence vs No history	1.16	0.14	0.17	0.8	0.38	-0.20	0.45
Correction history in the last 5 years	Custodial Sentence for non-violent crimes vs No history	1.28	0.24	0.25	0.9	0.34	-0.21	0.77
Correction history in the last 5 years	Custodial Sentence for violent crimes vs No history	1.09	0.09	0.21	0.2	0.66	-0.29	0.52
Caregiver's age	Under 20 vs 30_34/missing	1.91	0.65	0.17	13.9	0.00	0.25	0.92
Caregiver's age	20-24 vs 30-34	1.04	0.04	0.14	0.1	0.77	-0.23	0.30
Caregiver's age	25-29 vs 30-34	0.72	-0.33	0.15	4.7	0.03	-0.61	-0.02
Caregiver's age	35-39 vs 30-34	0.72	-0.32	0.22	2.1	0.15	-0.73	0.13
Caregiver's age	40 plus vs 30-34	1.25	0.23	0.30	0.6	0.45	-0.38	0.80
Time on benefit in last 5 years	More than 80% vs no time	2.28	0.82	0.14	35.4	<.0001	0.59	1.13
Time on benefit in last 5 years	20<-80% vs no time	1.71	0.54	0.13	18.6	<.0001	0.33	0.82
Time on benefit in last 5 years	Up to 20% vs no time	0.94	-0.06	0.17	0.1	0.74	-0.40	0.27
Mental health in last 5 years	Substance abuse vs No known issues	1.22	0.20	0.25	0.6	0.42	0.36	0.69
Mental health in last 5 years	Persistent substance abuse issues vs No known issues	1.41	0.34	0.33	1.1	0.30	-0.05	0.35
Mental health in last 5 years	Mental health issues other than substance abuse vs No known issues	0.79	-0.23	0.20	1.4	0.24	0.00	0.57
Mental health in last 5 years	Persistent mental health issues other than substance abuse vs No known issues	1.13	0.12	0.24	0.3	0.61	-0.53	0.15
Caregiver with care and protection history	Yes vs No	1.65	0.50	0.09	34.7	<.0001	0.27	1.06
Single parent	Yes vs Not single or unknown	1.15	0.14	0.10	1.9	0.16	-0.06	0.34
Single parent	Single and no father listed vs not single or unknown	1.35	0.30	0.14	4.4	0.04	0.02	0.58
Family violence	Events in ne of the last 12 months vs no events	0.85	-0.17	0.17	0.9	0.33	-0.51	0.17
Family violence	Events in more than one of the last 12 months vs no events	1.97	0.68	0.20	11.4	0.00	0.28	1.07

(h) Model based on benefit, care and protection, Corrections and Health data

Variable	Categories	Odds Ratio	Estimate	SE	Chi-Square	Sign.	95% CI Lower	95% CI Upper
Caregiver with care and protection history	Yes vs No	1.59	0.46	0.09	27.2	<.0001	0.32	0.66
Time on benefit in last 5 years	More than 80% vs no time	2.26	0.82	0.17	24.1	<.0001	0.55	1.19
Time on benefit in last 5 years	20<-80% vs no time	1.94	0.66	0.15	19.2	<.0001	0.39	0.98
Time on benefit in last 5 years	Up to 20% vs no time	1.13	0.12	0.21	0.3	0.58	-0.32	0.51
Caregiver with care and protection history	Yes vs No	1.66	0.51	0.10	28.2	<.0001	0.36	0.73
Address changes in last year	No address changes vs Missing	0.76	-0.28	0.16	3.2	0.08	-0.58	0.03
Address changes in last year	1 or 2 address changes vs Missing	1.10	0.09	0.16	0.3	0.56	-0.19	0.41
Address changes in last year	3 plus address changes vs Missing	1.92	0.65	0.24	7.2	0.01	0.17	1.11
Mental health in last 5 years	Substance abuse vs No known issues	1.32	0.28	0.22	1.6	0.21	0.20	0.55
Mental health in last 5 years	Persistent substance abuse issues vs No known issues	1.62	0.48	0.35	1.9	0.16	-0.20	1.16
Mental health in last 5 years	Mental health issues other than substance abuse vs No known issues	0.83	-0.19	0.19	1.0	0.32	-0.56	0.18
Mental health in last 5 years	Persistent mental health issues other than substance abuse vs No known issues	0.89	-0.12	0.24	0.2	0.62	-0.59	0.35
Mother is smoker	Yes vs No	1.41	0.35	0.09	14.9	0.00	0.17	0.52

(i) Model where no other children with history of contact with CYF

Variable	Categories	Odds Ratio	Estimate	SE	Chi-Square	Sign.	95% CI Lower	95% CI Upper
Benefit address changes in last year	No address changes vs Missing/no benefit in last year	0.73	-0.32	0.10	10.6	0.00	-0.51	-0.13
Benefit address changes in last year	1 or 2 address changes vs Missing/no benefit in last year	0.75	-0.29	0.10	8.8	0.00	-0.48	-0.09
Benefit address changes in last year	3 plus address changes vs Missing/no benefit in last year	1.49	0.40	0.12	10.4	0.00	0.15	0.64
Benefit caregiver is not birth registration parent	Yes vs No	1.73	0.55	0.20	7.4	0.01	0.16	0.94
Benefit caregiver is not birth registration parent	No birth registration vs No	1.00	-0.00	0.12	0.0	1.00	-0.25	0.23
Caregiver with care and protection history	Yes vs No	1.44	0.37	0.06	44.3	<.0001	0.32	0.52
Correction history in the last 5 years	Non-custodial sentence vs No history	0.80	-0.23	0.12	3.9	0.05	-0.43	0.02
Correction history in the last 5 years	Custodial Sentence for non-violent crimes vs No history	1.72	0.54	0.16	11.0	0.00	0.21	0.84
Correction history in the last 5 years	Custodial Sentence for violent crimes vs No history	1.58	0.46	0.14	11.0	0.00	0.18	0.72
Caregiver's age	Under 20 vs 30_34/missing	1.71	0.53	0.11	23.7	<.0001	0.33	0.76
Caregiver's age	20-24 vs 30-34	1.06	0.06	0.09	0.5	0.50	-0.12	0.23
Caregiver's age	25-29 vs 30-34	0.78	-0.25	0.10	6.3	0.01	-0.44	-0.06
Caregiver's age	35-39 vs 30-34	0.83	-0.19	0.15	1.6	0.20	-0.48	0.11
Caregiver's age	40 plus vs 30-34	1.09	0.09	0.20	0.2	0.66	-0.29	0.47
Behavioural or relationship difficulties as a child	Yes vs No	1.15	0.14	0.06	5.0	0.03	0.72	1.04
Time on benefit in last 5 years	More than 80% vs no time	2.41	0.88	0.08	111.7	<.0001	0.40	0.69
Time on benefit in last 5 years	20<-80% vs no time	1.71	0.53	0.07	51.5	<.0001	-0.29	0.07
Time on benefit in last 5 years	Up to 20% vs no time	0.91	-0.09	0.09	1.0	0.32	0.01	0.64
Mental health in last 5 years	Substance abuse vs No known issues	1.38	0.32	0.16	3.8	0.05	-0.00	0.87
Mental health in last 5 years	Persistent substance abuse issues vs No known issues	1.51	0.41	0.23	3.4	0.07	-0.24	0.28
Mental health in last 5 years	Mental health issues other than substance abuse vs No known issues	1.03	0.03	0.13	0.1	0.82	-0.02	0.57
Mental health in last 5 years	Persistent mental health issues other than substance abuse vs No known issues	1.30	0.26	0.15	3.0	0.08	0.14	0.39
Single parent	Yes vs Not single or unknown	1.31	0.27	0.06	17.3	<.0001	-0.02	0.35
Single parent	Single and no father listed vs not single or unknown	1.18	0.17	0.09	3.2	0.08	0.17	0.45
Low birth weight or pre-term	Yes vs No	1.37	0.32	0.07	19.3	<.0001	-0.02	0.26
Parenting demands	High parenting demands vs no high parenting demands	1.12	0.12	0.07	2.6	0.11	0.01	0.27
Parenting demands	No other children vs no parenting demands	1.14	0.13	0.07	4.0	0.05	0.52	1.55
Family violence	Events in ne of the last 12 months vs no events	2.83	1.04	0.26	15.7	<.0001	-0.90	0.37
Family violence	Events in more than one of the last 12 months vs no events	0.79	-0.23	0.32	0.5	0.47	-0.87	0.40
CYF service centre	Parameter estimates available upon request							

(j) Model where other children with history of contact with CYF

Variable	Categories	Odds Ratio	Estimate	SE	Chi-Square	Sign.	95% CI Lower	95% CI Upper
Caregiver with care and protection history	Yes vs No	1.35	0.30	0.05	37.6	<.0001	-1.20	0.05
Family violence	Events in one of the last 12 months vs no events	1.02	0.02	0.09	0.1	0.83	0.20	0.86
Family violence	Events in more than one of the last 12 months vs no events	1.51	0.41	0.10	18.6	<.0001	0.17	0.37

(k) Model for Maori ethnic group

Variable	Categories	Odds Ratio	Estimate	SE	Chi-Square	Sign.	95% CI Lower	95% CI Upper
Benefit address changes in last year	No address changes vs Missing/no benefit in last year	0.81	-0.21	0.10	4.5	0.03	0.03	0.69
Benefit address changes in last year	1 or 2 address changes vs Missing/no benefit in last year	0.95	-0.05	0.09	0.4	0.55	-0.00	0.40
Benefit address changes in last year	3 plus address changes vs Missing/no benefit in last year	1.08	0.07	0.12	0.4	0.55	0.19	0.38
Benefit caregiver is not birth registration parent	Yes vs No	1.52	0.42	0.17	6.1	0.01	-0.12	0.26
Benefit caregiver is not birth registration parent	No birth registration vs No	1.31	0.27	0.11	5.8	0.02	-0.11	0.43
Caregiver with care and protection history	Yes vs No	1.31	0.27	0.05	33.7	<.0001	-0.15	0.33
Correction history in the last 5 years	Non-custodial sentence vs No history	1.07	0.07	0.10	0.5	0.47	0.27	0.70
Correction history in the last 5 years	Custodial Sentence for non-violent crimes vs No history	1.19	0.17	0.14	1.5	0.22	-0.20	0.14
Correction history in the last 5 years	Custodial Sentence for violent crimes vs No history	1.09	0.09	0.12	0.5	0.47	-0.17	0.19
Caregiver's age	Under 20 vs 30_34/missing	1.59	0.46	0.11	18.2	<.0001	-0.47	0.12
Caregiver's age	20-24 vs 30-34	0.95	-0.05	0.09	0.4	0.53	-0.39	0.37
Caregiver's age	25-29 vs 30-34	0.99	-0.01	0.09	0.0	0.93	0.46	0.82
Caregiver's age	35-39 vs 30-34	0.85	-0.17	0.15	1.2	0.27	0.34	0.66
Caregiver's age	40 plus vs 30-34	1.05	0.05	0.20	0.1	0.80	-0.32	0.12
Time on benefit in last 5 years	More than 80% vs no time	1.88	0.63	0.09	48.0	<.0001	-0.48	0.15
Time on benefit in last 5 years	20<-80% vs no time	1.63	0.49	0.08	36.0	<.0001	-0.18	0.62
Time on benefit in last 5 years	Up to 20% vs no time	0.92	-0.09	0.11	0.6	0.45	-0.43	0.08
Mental health in last 5 years	Substance abuse vs No known issues	0.84	-0.17	0.16	1.1	0.29	0.25	0.84
Mental health in last 5 years	Persistent substance abuse issues vs No known issues	1.22	0.20	0.20	0.9	0.33	0.46	0.67
Mental health in last 5 years	Mental health issues other than substance abuse vs No known issues	0.85	-0.17	0.13	1.6	0.20	0.08	0.33
Mental health in last 5 years	Persistent mental health issues other than substance abuse vs No known issues	1.74	0.55	0.15	13.1	0.00	-0.14	0.21
Caregiver with care and protection history	Yes vs No	1.75	0.56	0.05	110.0	<.0001	-0.14	0.11
Single parent	Yes vs Not single or unknown	1.23	0.21	0.06	11.4	0.00	0.09	0.35
Single parent	Single and no father listed vs not single or unknown	1.03	0.03	0.09	0.1	0.73	-0.09	0.35
Parenting demands	High parenting demands vs no high parenting demands	0.99	-0.01	0.06	0.0	0.88	0.07	0.58
Parenting demands	No other children vs no parenting demands	1.23	0.21	0.07	9.9	0.00	0.08	0.34
Family violence	Events in one of the last 12 months vs no events	1.13	0.12	0.11	1.1	0.28	-0.10	0.34
Family violence	Events in more than one of the last 12 months vs no events	1.38	0.32	0.13	6.1	0.01	0.07	0.58

(k) Model for Maori ethnic group-continued

Variable	Categories	Odds Ratio	Estimate	SE	Chi-Square	Sign.	95% CI Lower	95% CI Upper
CYF Service centre	Blenheim vs Whangarei	2.02	0.70	0.46	2.3	0.13	-0.19	1.62
CYF Service centre	CYF Missing vs Whangarei	1.12	0.11	0.19	0.4	0.55	-0.24	0.50
CYF Service centre	Christchurch City vs Whangarei	1.29	0.25	0.31	0.7	0.41	-0.34	0.88
CYF Service centre	Clendon vs Whangarei	1.53	0.43	0.26	2.6	0.11	-0.10	0.94
CYF Service centre	Far North vs Whangarei	0.67	-0.39	0.27	2.2	0.14	-0.92	0.13
CYF Service centre	Gisborne vs Whangarei	1.09	0.09	0.21	0.2	0.68	-0.33	0.50
CYF Service centre	Grey Lynn vs Whangarei	1.50	0.40	0.30	1.8	0.18	-0.17	1.00
CYF Service centre	Hastings vs Whangarei	0.53	-0.63	0.26	6.1	0.01	-1.11	-0.11
CYF Service centre	Hauraki vs Whangarei	0.95	-0.05	0.34	0.0	0.88	-0.69	0.63
CYF Service centre	Hutt vs Whangarei	1.77	0.57	0.21	7.7	0.01	0.17	0.98
CYF Service centre	Manawatu vs Whangarei	0.62	-0.48	0.22	5.0	0.03	-0.90	-0.05
CYF Service centre	Mangere vs Whangarei	1.45	0.37	0.25	2.3	0.13	-0.13	0.83
CYF Service centre	Manurewa vs Whangarei	0.95	-0.05	0.23	0.1	0.82	-0.49	0.39
CYF Service centre	Napier vs Whangarei	0.46	-0.78	0.33	5.4	0.02	-1.40	-0.09
CYF Service centre	Nelson vs Whangarei	2.05	0.72	0.36	3.9	0.05	0.03	1.45
CYF Service centre	Onehunga vs Whangarei	1.91	0.65	0.41	2.5	0.12	-0.16	1.46
CYF Service centre	Orewa vs Whangarei	0.99	-0.01	0.41	0.0	0.99	-0.82	0.77
CYF Service centre	Otago Urban vs Whangarei	0.41	-0.90	0.60	2.3	0.13	-2.09	0.25
CYF Service centre	Otahuhu vs Whangarei	0.73	-0.31	0.33	0.9	0.35	-0.98	0.32
CYF Service centre	Otara vs Whangarei	1.45	0.37	0.30	1.5	0.21	-0.22	0.95
CYF Service centre	Panmure vs Whangarei	1.57	0.45	0.29	2.3	0.13	-0.11	1.03
CYF Service centre	Papakura vs Whangarei	1.66	0.51	0.23	4.8	0.03	0.04	0.94
CYF Service centre	Papanui vs Whangarei	0.96	-0.04	0.33	0.0	0.90	-0.68	0.63
CYF Service centre	Porirua vs Whangarei	0.35	-1.05	0.35	8.8	0.00	-1.75	-0.36
CYF Service centre	Pukekohe vs Whangarei	1.40	0.33	0.31	1.2	0.28	-0.29	0.93
CYF Service centre	Rotorua vs Whangarei	0.87	-0.14	0.20	0.5	0.48	-0.52	0.25
CYF Service centre	South Canterbury vs Whangarei	0.61	-0.49	0.41	1.5	0.23	-1.25	0.34
CYF Service centre	Southern Rural vs Whangarei	0.80	-0.23	0.54	0.2	0.67	-1.24	0.85
CYF Service centre	Southland vs Whangarei	0.81	-0.22	0.37	0.3	0.56	-0.96	0.50
CYF Service centre	Sydenham vs Whangarei	1.03	0.03	0.30	0.0	0.92	-0.52	0.64
CYF Service centre	Takapuna vs Whangarei	0.88	-0.12	0.35	0.1	0.72	-0.83	0.56
CYF Service centre	Taranaki vs Whangarei	0.48	-0.74	0.32	5.5	0.02	-1.38	-0.14
CYF Service centre	Tauranga vs Whangarei	0.88	-0.13	0.21	0.4	0.55	-0.53	0.30
CYF Service centre	Waikato East vs Whangarei	0.81	-0.21	0.22	0.9	0.34	-0.64	0.22
CYF Service centre	Waikato West vs Whangarei	0.78	-0.25	0.18	1.9	0.16	-0.60	0.10
CYF Service centre	Wairarapa vs Whangarei	0.84	-0.17	0.37	0.2	0.64	-0.90	0.55
CYF Service centre	Waitakere vs Whangarei	1.04	0.04	0.25	0.0	0.86	-0.47	0.52
CYF Service centre	Wellington vs Whangarei	0.96	-0.04	0.42	0.0	0.93	-0.83	0.79
CYF Service centre	West Coast vs Whangarei	2.88	1.06	0.52	4.2	0.04	-0.08	1.98
CYF Service centre	Westgate vs Whangarei	1.68	0.52	0.21	6.0	0.01	0.11	0.93
CYF Service centre	Whakatane vs Whangarei	2.15	0.76	0.18	18.7	<.0001	0.40	1.10
CYF Service centre	Whanganui vs Whangarei	0.42	-0.86	0.29	8.6	0.00	-1.42	-0.27

(I) Model for “Non-Maori and ethnic group not recorded”

Variable	Categories	Odds Ratio	Estimate	SE	Chi-Square	Sign.	95% CI Lower	95% CI Upper
Caregiver with care and protection history	Yes vs No	1.45	0.37	0.08	21.7	<.0001	0.23	0.54
Correction history in the last 5 years	Non-custodial sentence vs No history	1.01	0.01	0.18	0.0	0.97	-0.38	0.34
Correction history in the last 5 years	Custodial Sentence for non-violent crimes vs No history	0.90	-0.11	0.30	0.1	0.71	-0.70	0.48
Correction history in the last 5 years	Custodial Sentence for violent crimes vs No history	2.01	0.70	0.22	9.9	0.00	0.25	1.12
Caregiver's age	Under 20 vs 30_34/missing	2.71	1.00	0.15	46.2	<.0001	0.57	1.18
Caregiver's age	20-24 vs 30-34	1.06	0.06	0.11	0.3	0.61	-0.14	0.29
Caregiver's age	25-29 vs 30-34	0.46	-0.77	0.13	34.8	<.0001	-1.01	-0.49
Caregiver's age	35-39 vs 30-34	0.74	-0.30	0.17	3.2	0.08	-0.61	0.07
Caregiver's age	40 plus vs 30-34	1.34	0.29	0.23	1.6	0.21	-0.14	0.76
Behavioural or relationship difficulties as a child	Yes vs No	1.34	0.29	0.09	10.5	0.00	0.11	0.47
Time on benefit in last 5 years	More than 80% vs no time	2.00	0.69	0.11	38.4	<.0001	0.52	0.97
Time on benefit in last 5 years	20<-80% vs no time	1.78	0.58	0.10	34.2	<.0001	0.41	0.81
Time on benefit in last 5 years	Up to 20% vs no time	1.13	0.12	0.12	1.0	0.31	-0.13	0.34
Mental health in last 5 years	Substance abuse vs No known issues	1.93	0.66	0.21	10.1	0.00	0.24	1.05
Mental health in last 5 years	Persistent substance abuse issues vs No known issues	4.37	1.48	0.29	25.4	<.0001	0.96	2.13
Mental health in last 5 years	Mental health issues other than substance abuse vs No known issues	0.56	-0.58	0.16	12.6	0.00	-0.96	-0.31
Mental health in last 5 years	Persistent mental health issues other than substance abuse vs No known issues	0.65	-0.43	0.19	5.0	0.03	-0.85	-0.08
Caregiver with care and protection history	Yes vs No	1.89	0.64	0.08	69.1	<.0001	0.54	0.86
Single parent	Yes vs Not single or unknown	1.00	0.00	0.09	0.0	0.97	-0.18	0.17
Single parent	Single and no father listed vs not single or unknown	1.69	0.52	0.13	16.8	<.0001	0.24	0.74
Family violence	Events in one of the last 12 months vs no events	1.47	0.38	0.21	3.5	0.06	-0.16	0.20
Family violence	Events in more than one of the last 12 months vs no events	2.79	1.03	0.27	14.6	0.00	0.06	0.42

(I) Model for “Non-Maori and ethnic group not recorded” - continued

Variable	Categories	Odds Ratio	Estimate	SE	Chi-Square	Sign.	95% CI Lower	95% CI Upper
CYF Service centre	Blenheim vs Whangarei	0.20	-1.63	1.01	2.6	0.11	-3.60	0.35
CYF Service centre	CYF Missing vs Whangarei	0.76	-0.28	0.26	1.2	0.28	-0.76	0.25
CYF Service centre	Christchurch City vs Whangarei	0.36	-1.02	0.30	11.4	0.00	-1.58	-0.40
CYF Service centre	Clendon vs Whangarei	2.35	0.85	0.37	5.4	0.02	0.20	1.65
CYF Service centre	Far North vs Whangarei	6.06	1.80	0.40	20.5	<.0001	1.00	2.57
CYF Service centre	Gisborne vs Whangarei	0.16	-1.85	0.60	9.4	0.00	-3.07	-0.71
CYF Service centre	Grey Lynn vs Whangarei	1.57	0.45	0.27	2.8	0.10	-0.11	0.96
CYF Service centre	Hastings vs Whangarei	0.44	-0.82	0.59	1.9	0.16	-1.88	0.45
CYF Service centre	Hauraki vs Whangarei	0.55	-0.59	0.49	1.4	0.23	-1.69	0.27
CYF Service centre	Hutt vs Whangarei	1.64	0.50	0.25	4.0	0.05	0.04	1.03
CYF Service centre	Manawatu vs Whangarei	0.75	-0.28	0.27	1.1	0.30	-0.77	0.29
CYF Service centre	Mangere vs Whangarei	0.95	-0.05	0.26	0.0	0.83	-0.60	0.41
CYF Service centre	Manurewa vs Whangarei	6.04	1.80	0.29	37.4	<.0001	1.28	2.44
CYF Service centre	Napier vs Whangarei	0.94	-0.06	0.55	0.0	0.92	-0.93	1.26
CYF Service centre	Nelson vs Whangarei	0.29	-1.25	0.41	9.1	0.00	-2.10	-0.49
CYF Service centre	Onehunga vs Whangarei	0.68	-0.39	0.38	1.1	0.30	-1.05	0.40
CYF Service centre	Orewa vs Whangarei	1.77	0.57	0.34	2.7	0.10	-0.15	1.21
CYF Service centre	Otago Urban vs Whangarei	0.72	-0.33	0.37	0.8	0.37	-1.08	0.38
CYF Service centre	Otahuhu vs Whangarei	0.98	-0.02	0.31	0.0	0.95	-0.64	0.57
CYF Service centre	Otara vs Whangarei	1.82	0.60	0.30	3.9	0.05	0.05	1.24
CYF Service centre	Panmure vs Whangarei	1.26	0.23	0.37	0.4	0.54	-0.48	0.98
CYF Service centre	Papakura vs Whangarei	0.82	-0.19	0.39	0.3	0.62	-0.96	0.58
CYF Service centre	Papanui vs Whangarei	0.59	-0.53	0.30	3.3	0.07	-1.14	0.03
CYF Service centre	Porirua vs Whangarei	0.82	-0.20	0.39	0.3	0.61	-0.91	0.63
CYF Service centre	Pukekohe vs Whangarei	0.21	-1.54	1.00	2.4	0.12	-3.58	0.34
CYF Service centre	Rotorua vs Whangarei	1.68	0.52	0.44	1.4	0.24	-0.39	1.32
CYF Service centre	South Canterbury vs Whangarei	0.85	-0.16	0.42	0.1	0.71	-0.98	0.69
CYF Service centre	Southern Rural vs Whangarei	2.68	0.99	0.41	5.9	0.02	0.29	1.89
CYF Service centre	Southland vs Whangarei	0.39	-0.94	0.47	4.0	0.05	-1.99	-0.13
CYF Service centre	Sydenham vs Whangarei	0.25	-1.38	0.35	15.7	<.0001	-2.14	-0.75
CYF Service centre	Takapuna vs Whangarei	3.76	1.32	0.30	19.6	<.0001	0.75	1.93
CYF Service centre	Taranaki vs Whangarei	1.33	0.29	0.41	0.5	0.49	-0.61	1.04
CYF Service centre	Tauranga vs Whangarei	2.15	0.77	0.27	8.3	0.00	0.25	1.30
CYF Service centre	Waikato East vs Whangarei	1.11	0.11	0.31	0.1	0.73	-0.53	0.67
CYF Service centre	Waikato West vs Whangarei	1.06	0.06	0.27	0.1	0.83	-0.51	0.57
CYF Service centre	Wairarapa vs Whangarei	0.56	-0.59	0.54	1.2	0.28	-1.62	0.48
CYF Service centre	Waitakere vs Whangarei	0.83	-0.19	0.27	0.5	0.48	-0.70	0.34
CYF Service centre	Wellington vs Whangarei	0.55	-0.60	0.58	1.1	0.30	-1.83	0.46
CYF Service centre	West Coast vs Whangarei	10.22	2.32	0.57	16.9	<.0001	1.26	3.53
CYF Service centre	Westgate vs Whangarei	0.98	-0.02	0.28	0.0	0.95	-0.55	0.52
CYF Service centre	Whakatane vs Whangarei	4.99	1.61	0.39	17.2	<.0001	0.79	2.32
CYF Service centre	Whanganui vs Whangarei	0.39	-0.94	0.54	3.0	0.08	-2.04	0.12

(m) Model for beneficiaries using link 3c

Variable	Categories	Odds Ratio	Estimate	SE	Chi-Square	Sign.	95% CI Lower	95% CI Upper
Benefit address changes in last year	No address changes vs Missing/no benefit in last year	0.74	-0.30	0.08	14.8	0.00	-0.46	-0.15
Benefit address changes in last year	1 or 2 address changes vs Missing/no benefit in last year	0.84	-0.17	0.07	5.6	0.02	-0.31	-0.03
Benefit address changes in last year	3 plus address changes vs Missing/no benefit in last year	1.29	0.26	0.09	7.7	0.01	0.08	0.44
Benefit caregiver is not birth registration parent	Yes vs No	1.46	0.38	0.13	7.9	0.01	0.12	0.65
Benefit caregiver is not birth registration parent	No birth registration vs No	1.05	0.05	0.09	0.3	0.57	-0.14	0.20
Caregiver with care and protection history	Yes vs No	1.29	0.25	0.04	32.8	<.0001	0.17	0.34
Correction history in the last 5 years	Non-custodial sentence vs No history	1.03	0.03	0.08	0.1	0.72	-0.14	0.20
Correction history in the last 5 years	Custodial Sentence for non-violent crimes vs No history	1.34	0.29	0.12	5.8	0.02	0.06	0.54
Correction history in the last 5 years	Custodial Sentence for violent crimes vs No history	1.02	0.02	0.11	0.0	0.88	-0.20	0.23
Caregiver's age	Under 20 vs 30_34/missing	1.46	0.38	0.09	18.1	<.0001	0.20	0.55
Caregiver's age	20-24 vs 30-34	0.95	-0.05	0.07	0.5	0.50	-0.19	0.09
Caregiver's age	25-29 vs 30-34	0.75	-0.28	0.08	12.5	0.00	-0.45	-0.13
Caregiver's age	35-39 vs 30-34	1.00	0.00	0.12	0.0	0.98	-0.23	0.24
Caregiver's age	40 plus vs 30-34	1.12	0.12	0.15	0.6	0.44	-0.17	0.42
Behavioural or relationship difficulties as a child	Yes vs No	1.23	0.20	0.05	17.2	<.0001	0.11	0.30
Time on benefit in last 5 years	More than 80% vs no time	1.61	0.48	0.09	25.9	<.0001	0.29	0.66
Time on benefit in last 5 years	20<-80% vs no time	1.39	0.33	0.09	13.4	0.00	0.15	0.51
Time on benefit in last 5 years	Up to 20% vs no time	0.80	-0.23	0.11	4.1	0.04	-0.45	-0.01
Mental health in last 5 years	Substance abuse vs No known issues	1.16	0.15	0.12	1.5	0.22	-0.09	0.39
Mental health in last 5 years	Persistent substance abuse issues vs No known issues	1.23	0.20	0.16	1.6	0.21	-0.11	0.52
Mental health in last 5 years	Mental health issues other than substance abuse vs No known issues	0.82	-0.20	0.10	4.2	0.04	-0.40	-0.01
Mental health in last 5 years	Persistent mental health issues other than substance abuse vs No known issues	1.25	0.23	0.11	4.0	0.04	0.01	0.45
Caregiver with care and protection history	Yes vs No	1.58	0.46	0.04	107.1	<.0001	0.37	0.54
Low birth weight or pre-term	Yes vs No	1.11	0.11	0.06	3.2	0.07	-0.03	0.18
Parenting demands	High parenting demands vs no high parenting demands	1.08	0.07	0.05	2.0	0.16	-0.00	0.22
Parenting demands	No other children vs no parenting demands	1.11	0.11	0.06	3.7	0.05	0.00	0.37
Family violence	Events in one of the last 12 months vs no events	1.20	0.18	0.09	3.9	0.05	0.08	0.49
Family violence	Events in more than one of the last 12 months vs no events	1.33	0.28	0.11	7.1	0.01	0.08	0.49
CYF service centre	Parameter estimates available upon request							

(n) Model for beneficiaries using link 4c

Variable	Categories	Odds Ratio	Estimate	SE	Chi-Square	Sign.	95% CI Lower	95% CI Upper
Benefit caregiver is not birth registration parent	Yes vs No	1.33	0.28	0.14	3.9	0.05	0.00	0.56
Benefit caregiver is not birth registration parent	No birth registration vs No	1.01	0.01	0.08	0.0	0.89	-0.15	0.18
Caregiver with care and protection history	Yes vs No	1.26	0.23	0.04	25.8	<.0001	0.14	0.32
Correction history in the last 5 years	Non-custodial sentence vs No history	1.10	0.09	0.09	1.2	0.28	-0.08	0.26
Correction history in the last 5 years	Custodial Sentence for non-violent crimes vs No history	1.21	0.19	0.12	2.4	0.12	-0.05	0.44
Correction history in the last 5 years	Custodial Sentence for violent crimes vs No history	1.03	0.03	0.11	0.1	0.80	-0.19	0.25
Caregiver's age	Under 20 vs 30_34/missing	1.59	0.46	0.09	26.4	<.0001	0.29	0.64
Caregiver's age	20-24 vs 30-34	1.03	0.03	0.07	0.2	0.70	-0.11	0.18
Caregiver's age	25-29 vs 30-34	0.72	-0.33	0.08	15.6	<.0001	-0.49	-0.16
Caregiver's age	35-39 vs 30-34	0.99	-0.01	0.12	0.0	0.93	-0.25	0.22
Caregiver's age	40 plus vs 30-34	0.85	-0.16	0.17	0.9	0.34	-0.49	0.15
Behavioural or relationship difficulties as a child	Yes vs No	1.18	0.17	0.05	11.3	0.00	0.07	0.27
Time on benefit in last 5 years	More than 80% vs no time	1.42	0.35	0.09	14.2	0.00	0.17	0.54
Time on benefit in last 5 years	20<-80% vs no time	1.31	0.27	0.09	9.6	0.00	0.10	0.44
Time on benefit in last 5 years	Up to 20% vs no time	0.82	-0.20	0.11	3.4	0.07	-0.42	0.01
Caregiver with care and protection history	Yes vs No	1.53	0.42	0.04	90.4	<.0001	0.34	0.52
Address changes in last year	No address changes vs Missing	0.80	-0.23	0.07	12.0	0.00	-0.35	-0.10
Address changes in last year	1 or 2 address changes vs Missing	0.97	-0.03	0.06	0.2	0.68	-0.15	0.10
Address changes in last year	3 plus address changes vs Missing	1.20	0.18	0.10	3.4	0.06	-0.01	0.38
Mental health in last 5 years	Substance abuse vs No known issues	1.31	0.27	0.09	9.2	0.00	0.10	0.45
Mental health in last 5 years	Persistent substance abuse issues vs No known issues	1.32	0.28	0.14	3.8	0.05	0.00	0.56
Mental health in last 5 years	Mental health issues other than substance abuse vs No known issues	0.88	-0.13	0.08	2.8	0.10	-0.28	0.02
Mental health in last 5 years	Persistent mental health issues other than substance abuse vs No known issues	1.06	0.05	0.10	0.3	0.58	-0.14	0.24
Low birth weight or pre-term	Yes vs No	1.14	0.13	0.05	6.6	0.01	0.03	0.23
Parenting demands	High parenting demands vs no high parenting demands	1.13	0.12	0.05	5.2	0.02	0.03	0.23
Parenting demands	No other children vs no parenting demands	1.04	0.04	0.06	0.4	0.51	-0.08	0.15
Family violence	Events in one of the last 12 months vs no events	1.08	0.08	0.09	0.7	0.40	-0.10	0.27
Family violence	Events in more than one of the last 12 months vs no events	1.37	0.32	0.09	12.6	0.00	0.14	0.49
Other children immunisation history	Yes vs No	1.08	0.08	0.06	1.8	0.18	0.13	0.28
Mother is smoker	1 vs 2	1.23	0.20	0.04	29.5	<.0001	0.13	0.28

(o) Model for non- beneficiaries using link 3c

Variable	Categories	Odds Ratio	Estimate	SE	Chi-Square	Sign.	95% CI Lower	95% CI Upper
Caregiver with care and protection history	Yes vs No	2.10	0.74	0.05	230.7	<.0001	-0.85	0.9985
Correction history in the last 5 years	Non-custodial sentence vs No history	0.45	-0.79	0.12	41.5	<.0001	-0.60	0.1403
Correction history in the last 5 years	Custodial Sentence for non-violent crimes vs No history	3.40	1.22	0.16	57.4	<.0001	-0.14	0.87
Correction history in the last 5 years	Custodial Sentence for violent crimes vs No history	1.45	0.38	0.15	6.3	0.01	-0.22	0.9752
Time on benefit in last 5 years	more than 80% vs no time	2.24	0.81	0.09	78.2	<.0001	-0.37	1.4915
Time on benefit in last 5 years	20<-80% vs no time	1.29	0.26	0.07	12.1	0.00	0.10	1.2003
Time on benefit in last 5 years	Up to 20% vs no time	0.99	-0.01	0.09	0.0	0.88	-0.34	0.6409
Mental health in last 5 years	Substance abuse vs No known issues	0.97	-0.03	0.19	0.0	0.89	-1.41	0.0372
Mental health in last 5 years	Persistent substance abuse issues vs No known issues	7.80	2.05	0.30	46.1	<.0001	-0.31	1.0608
Mental health in last 5 years	Mental health issues other than substance abuse vs No known issues	0.45	-0.79	0.15	28.5	<.0001	-0.58	0.4481
Mental health in last 5 years	Persistent mental health issues other than substance abuse vs No known issues	0.98	-0.02	0.22	0.0	0.92	-3.13	-1.666
Other children with care and protection history	Yes vs No	2.44	0.89	0.07	187.5	<.0001	-0.22	0.7708

(p) Model for non- beneficiaries using link 4c

Variable	Categories	Odds Ratio	Estimate	SE	Chi-Square	Sign.	95% CI Lower	95% CI Upper
Caregiver with care and protection history	Yes vs No	1.96	0.67	0.05	182.3	<.0001	-0.66	1.3209
Time on benefit in last 5 years	more than 80% vs no time	2.57	0.94	0.09	122.0	<.0001	0.15	0.7783
Time on benefit in last 5 years	20<-80% vs no time	1.50	0.40	0.07	34.0	<.0001	-0.85	0.3199
Time on benefit in last 5 years	Up to 20% vs no time	0.90	-0.10	0.08	1.5	0.23	-0.42	0.9391
Other children with care and protection history	Yes vs No	1.79	0.58	0.06	90.4	<.0001	0.05	2.0319

(q) Model to predict substantiated physical or sexual abuse or neglect (excluding emotional abuse findings)

Variable	Categories	Odds Ratio	Estimate	SE	Chi-Square	Sign.	95% CI Lower	95% CI Upper
Benefit caregiver is not birth registration parent	Yes vs No	1.53	0.42	0.21	4.2	0.04	-1.51	2.40
Benefit caregiver is not birth registration parent	No birth registration vs No	1.18	0.17	0.12	2.0	0.16	-0.82	0.12
Caregiver with care and protection history	Yes vs No	1.60	0.47	0.05	83.2	<.0001	-1.08	0.13
Time on benefit in last 5 years	More than 80% vs no time	2.63	0.97	0.10	101.0	<.0001	-0.20	1.15
Time on benefit in last 5 years	20<-80% vs no time	2.26	0.82	0.09	87.9	<.0001	0.18	1.35
Time on benefit in last 5 years	Up to 20% vs no time	0.74	-0.30	0.14	4.8	0.03	0.05	1.23
Other children with care and protection history	Yes vs No	2.35	0.85	0.06	220.4	<.0001	-0.46	0.88
Parenting demands	High parenting demands vs no high parenting demands	0.98	-0.03	0.07	0.1	0.72	-0.84	0.49
Parenting demands	No other children vs no parenting demands	1.56	0.45	0.08	34.4	<.0001	-1.27	0.30

(r) Model to predict investigations or Child and Family Assessments

Variable	Categories	Odds Ratio	Estimate	SE	Chi-Square	Sign.	95% CI Lower	95% CI Upper
Caregiver with care and protection history	Yes vs No	1.48	0.39	0.03	208.0	<.0001	-0.37	0.81
Caregiver's age	Under 20 vs 30_34/missing	1.77	0.57	0.06	96.3	<.0001	-0.26	0.18
Caregiver's age	20-24 vs 30-34	0.94	-0.06	0.05	1.8	0.18	-0.36	0.21
Caregiver's age	25-29 vs 30-34	0.76	-0.28	0.05	27.3	<.0001	-0.68	0.06
Caregiver's age	35-39 vs 30-34	0.80	-0.22	0.07	8.9	0.00	-0.74	-0.03
Caregiver's age	40 plus vs 30-34	1.14	0.13	0.10	1.7	0.19	-0.11	0.55
Time on benefit in last 5 years	More than 80% vs no time	1.92	0.65	0.05	197.9	<.0001	-0.26	0.33
Time on benefit in last 5 years	20<-80% vs no time	1.70	0.53	0.04	171.7	<.0001	-0.21	0.37
Time on benefit in last 5 years	Up to 20% vs no time	1.03	0.03	0.05	0.3	0.60	0.11	0.77
Mental health in last 5 years	Substance abuse vs No known issues	1.20	0.18	0.09	4.3	0.04	0.11	0.58
Mental health in last 5 years	Persistent substance abuse issues vs No known issues	1.87	0.63	0.11	30.1	<.0001	-0.58	-0.08
Mental health in last 5 years	Mental health issues other than substance abuse vs No known issues	0.90	-0.10	0.07	2.3	0.13	-0.29	0.25
Mental health in last 5 years	Persistent mental health issues other than substance abuse vs No known issues	0.98	-0.02	0.09	0.1	0.81	-0.10	0.41
Other children with care and protection history	Yes vs No	1.83	0.61	0.03	463.7	<.0001	-0.98	-0.17
Single parent	Yes vs Not single or unknown	1.26	0.23	0.04	40.7	<.0001	-0.72	0.13
Single parent	Single and no father listed vs not single or unknown	1.25	0.22	0.05	18.4	<.0001	-0.43	0.34

(s) Model to predict substantiated physical abuse or hospitalisation for maltreatment or marker injury hospitalisation

Variable	Categories	Odds Ratio	Estimate	SE	Chi-Square	Sign.	95% CI Lower	95% CI Upper
Benefit caregiver is not birth registration parent	Yes vs No	2.07	0.73	0.20	13.6	0.00	-1.70	1.05
Benefit caregiver is not birth registration parent	No birth registration vs No	0.90	-0.11	0.12	0.8	0.36	0.07	0.73
Caregiver with care and protection history	Yes vs No	1.44	0.37	0.06	40.8	<.0001	0.02	0.98
Caregiver's age	Under 20 vs 30_34/missing	1.70	0.53	0.11	21.5	<.0001	-0.82	0.69
Caregiver's age	20-24 vs 30-34	0.91	-0.09	0.10	0.8	0.36	-1.99	0.25
Caregiver's age	25-29 vs 30-34	1.00	-0.00	0.10	0.0	0.98	0.24	1.30
Caregiver's age	35-39 vs 30-34	1.17	0.16	0.13	1.6	0.21	0.22	1.11
Caregiver's age	40 plus vs 30-34	0.69	-0.38	0.22	2.9	0.09	-0.83	0.32
Time on benefit in last 5 years	More than 80% vs no time	1.66	0.51	0.09	33.1	<.0001	-1.43	0.55
Time on benefit in last 5 years	20<-80% vs no time	1.29	0.25	0.08	10.6	0.00	0.15	1.11
Time on benefit in last 5 years	Up to 20% vs no time	0.97	-0.03	0.10	0.1	0.75	-1.31	-0.10
Other children with care and protection history	Yes vs No	1.67	0.51	0.06	72.4	<.0001	-0.29	0.71

(t) Model to predict notifications (including Police FV notifications and contact records)

Variable	Categories	Odds Ratio	Estimate	SE	Chi-Square	Sign.	95% CI Lower	95% CI Upper
Benefit address changes in last year	No address changes vs Missing/no benefit in last year	0.78	-0.25	0.05	30.1	<.0001	-0.34	-0.16
Benefit address changes in last year	1 or 2 address changes vs Missing/no benefit in last year	0.95	-0.06	0.04	1.5	0.22	-0.14	0.03
Benefit address changes in last year	3 plus address changes vs Missing/no benefit in last year	1.32	0.28	0.07	17.7	<.0001	0.15	0.40
Benefit caregiver is not birth registration parent	Yes vs No	1.29	0.26	0.10	6.2	0.01	0.05	0.46
Benefit caregiver is not birth registration parent	No birth registration vs No	1.02	0.02	0.06	0.1	0.73	-0.10	0.14
Caregiver with care and protection history	Yes vs No	1.36	0.30	0.03	134.7	<.0001	0.25	0.36
Correction history in the last 5 years	Non-custodial sentence vs No history	1.06	0.06	0.06	1.2	0.27	-0.05	0.17
Correction history in the last 5 years	Custodial Sentence for non-violent crimes vs No history	1.28	0.25	0.08	9.0	0.00	0.09	0.41
Correction history in the last 5 years	Custodial Sentence for violent crimes vs No history	1.13	0.13	0.07	3.1	0.08	-0.01	0.27
Caregiver's age	Under 20 vs 30_34/missing	1.55	0.44	0.05	66.9	<.0001	0.33	0.54
Caregiver's age	20-24 vs 30-34	1.03	0.03	0.04	0.7	0.42	-0.05	0.11
Caregiver's age	25-29 vs 30-34	0.81	-0.21	0.04	23.3	<.0001	-0.30	-0.13
Caregiver's age	35-39 vs 30-34	0.90	-0.10	0.06	2.8	0.09	-0.22	0.02
Caregiver's age	40 plus vs 30-34	1.05	0.05	0.09	0.3	0.60	-0.13	0.22
Behavioural or relationship difficulties as a child	Yes vs No	1.11	0.10	0.03	9.7	0.00	0.04	0.17
Time on benefit in last 5 years	More than 80% vs no time	1.69	0.52	0.04	160.8	<.0001	0.44	0.60
Time on benefit in last 5 years	20<-80% vs no time	1.64	0.49	0.03	215.7	<.0001	0.43	0.56
Time on benefit in last 5 years	Up to 20% vs no time	1.12	0.11	0.04	7.4	0.01	0.03	0.19
Mental health in last 5 years	Substance abuse vs No known issues	1.26	0.23	0.08	7.3	0.01	0.06	0.39
Mental health in last 5 years	Persistent substance abuse issues vs No known issues	1.57	0.45	0.12	14.8	0.00	0.22	0.68
Mental health in last 5 years	Mental health issues other than substance abuse vs No known issues	0.92	-0.09	0.06	1.9	0.16	-0.21	0.04
Mental health in last 5 years	Persistent mental health issues other than substance abuse vs No known issues	1.01	0.01	0.08	0.0	0.89	-0.14	0.17
Other children with care and protection history	Yes vs No	1.60	0.47	0.03	277.6	<.0001	0.42	0.53
Single parent	Yes vs Not single or unknown	1.22	0.20	0.03	39.4	<.0001	0.14	0.26
Single parent	Single and no father listed vs not single or unknown	1.17	0.16	0.05	12.0	0.00	0.07	0.25
Low birth weight or pre-term	Yes vs No	1.11	0.10	0.03	8.9	0.00	0.03	0.17
Parenting demands	High parenting demands vs no high parenting demands	1.08	0.08	0.03	6.5	0.01	0.02	0.14
Parenting demands	No other children vs no parenting demands	1.10	0.10	0.03	9.8	0.00	0.04	0.16
Family violence	Events in one of the last 12 months vs no events	1.09	0.08	0.07	1.3	0.25	-0.06	0.22
Family violence	Events in more than one of the last 12 months vs no events	1.44	0.37	0.09	18.1	<.0001	0.20	0.53

(u) Model with no local variables

Variable	Categories	Odds Ratio	Estimate	SE	Chi-Square	Sign.	95% CI Lower	95% CI Upper
Benefit address changes in last year	No address changes vs Missing/no benefit in last year	0.83	-0.19	0.07	6.9	0.01	-0.33	-0.05
Benefit address changes in last year	1 or 2 address changes vs Missing/no benefit in last year	0.77	-0.26	0.07	13.6	0.00	-0.40	-0.12
Benefit address changes in last year	3 plus address changes vs Missing/no benefit in last year	1.37	0.31	0.10	10.7	0.00	0.13	0.50
Benefit caregiver is not birth registration parent	Yes vs No	1.96	0.67	0.13	25.5	<.0001	0.41	0.94
Benefit caregiver is not birth registration parent	No birth registration vs No	1.00	0.00	0.08	0.0	0.98	-0.16	0.17
Caregiver with care and protection history	Yes vs No	1.30	0.26	0.04	37.9	<.0001	0.18	0.34
Correction history in the last 5 years	Non-custodial sentence vs No history	1.11	0.10	0.08	1.6	0.21	-0.06	0.26
Correction history in the last 5 years	Custodial Sentence for non-violent crimes vs No history	0.90	-0.11	0.12	0.8	0.37	-0.34	0.12
Correction history in the last 5 years	Custodial Sentence for violent crimes vs No history	1.42	0.35	0.10	11.3	0.00	0.14	0.55
Caregiver's age	Under 20 vs 30_34/missing	1.48	0.40	0.09	21.4	<.0001	0.23	0.56
Caregiver's age	20-24 vs 30-34	0.88	-0.13	0.07	4.0	0.05	-0.26	-0.00
Caregiver's age	25-29 vs 30-34	0.88	-0.13	0.07	3.2	0.07	-0.26	0.01
Caregiver's age	35-39 vs 30-34	0.97	-0.03	0.10	0.1	0.74	-0.24	0.17
Caregiver's age	40 plus vs 30-34	1.20	0.18	0.14	1.6	0.20	-0.10	0.46
Behavioural or relationship difficulties as a child	Yes vs No	1.21	0.19	0.05	17.6	<.0001	0.10	0.28
Time on benefit in last 5 years	More than 80% vs no time	2.03	0.71	0.07	109.5	<.0001	0.58	0.84
Time on benefit in last 5 years	20<-80% vs no time	1.84	0.61	0.06	109.1	<.0001	0.50	0.72
Time on benefit in last 5 years	Up to 20% vs no time	1.10	0.10	0.08	1.5	0.22	-0.06	0.25
Mental health in last 5 years	Substance abuse vs No known issues	1.24	0.21	0.12	3.2	0.07	-0.02	0.45
Mental health in last 5 years	Persistent substance abuse issues vs No known issues	1.61	0.48	0.15	10.0	0.00	0.18	0.77
Mental health in last 5 years	Mental health issues other than substance abuse vs No known issues	0.75	-0.29	0.09	10.3	0.00	-0.47	-0.11
Mental health in last 5 years	Persistent mental health issues other than substance abuse vs No known issues	1.11	0.10	0.11	0.9	0.36	-0.12	0.32
Other children with care and protection history	Yes vs No	1.74	0.56	0.04	189.4	<.0001	0.48	0.64
Single parent	Yes vs Not single or unknown	1.16	0.15	0.05	9.1	0.00	0.05	0.24
Single parent	Single and no father listed vs not single or unknown	1.17	0.16	0.07	5.1	0.02	0.02	0.30
Parenting demands	High parenting demands vs no high parenting demands	1.14	0.13	0.05	7.2	0.01	0.04	0.22
Parenting demands	No other children vs no parenting demands	1.04	0.04	0.05	0.7	0.41	-0.06	0.14
Family violence	Events in one of the last 12 months vs no events	1.03	0.03	0.09	0.2	0.70	-0.14	0.21
Family violence	Events in more than one of the last 12 months vs no events	1.67	0.51	0.10	25.2	<.0001	0.31	0.71

(v) Model with aggregated local level variables

Variable	Categories	Odds Ratio	Estimate	SE	Chi-Square	Sign.	95% CI Lower	95% CI Upper
Benefit address changes in last year	No address changes vs Missing/no benefit in last year	0.83	-0.19	0.07	6.8	0.01	-0.33	-0.05
Benefit address changes in last year	1 or 2 address changes vs Missing/no benefit in last year	0.77	-0.26	0.07	13.5	0.00	-0.40	-0.12
Benefit address changes in last year	3 plus address changes vs Missing/no benefit in last year	1.36	0.31	0.10	10.2	0.00	0.12	0.50
Benefit caregiver is not birth registration parent	Yes vs No	1.96	0.67	0.13	25.3	<.0001	0.41	0.93
Benefit caregiver is not birth registration parent	No birth registration vs No	1.00	-0.00	0.08	0.0	1.00	-0.17	0.16
Caregiver with care and protection history	Yes vs No	1.29	0.26	0.04	37.2	<.0001	0.18	0.34
Correction history in the last 5 years	Non-custodial sentence vs No history	1.09	0.09	0.08	1.2	0.28	-0.07	0.25
Correction history in the last 5 years	Custodial Sentence for non-violent crimes vs No history	0.90	-0.10	0.12	0.7	0.40	-0.33	0.13
Correction history in the last 5 years	Custodial Sentence for violent crimes vs No history	1.41	0.35	0.10	11.2	0.00	0.14	0.55
Caregiver's age	Under 20 vs 30_34/missing	1.48	0.39	0.09	21.2	<.0001	0.23	0.56
Caregiver's age	20-24 vs 30-34	0.88	-0.13	0.07	4.2	0.04	-0.26	-0.01
Caregiver's age	25-29 vs 30-34	0.88	-0.13	0.07	3.3	0.07	-0.27	0.01
Caregiver's age	35-39 vs 30-34	0.96	-0.04	0.10	0.1	0.71	-0.24	0.16
Caregiver's age	40 plus vs 30-34	1.22	0.20	0.14	2.0	0.16	-0.08	0.48
Behavioural or relationship difficulties as a child	Yes vs No	1.21	0.19	0.05	17.1	<.0001	0.10	0.28
Time on benefit in last 5 years	More than 80% vs no time	2.03	0.71	0.07	109.8	<.0001	0.58	0.84
Time on benefit in last 5 years	20<-80% vs no time	1.83	0.60	0.06	106.7	<.0001	0.49	0.72
Time on benefit in last 5 years	Up to 20% vs no time	1.10	0.10	0.08	1.6	0.21	-0.05	0.25
Mental health in last 5 years	Substance abuse vs No known issues	1.22	0.20	0.12	2.9	0.09	-0.03	0.44
Mental health in last 5 years	Persistent substance abuse issues vs No known issues	1.63	0.49	0.15	10.5	0.00	0.19	0.78
Mental health in last 5 years	Mental health issues other than substance abuse vs No known issues	0.75	-0.29	0.09	10.1	0.00	-0.47	-0.11
Mental health in last 5 years	Persistent mental health issues other than substance abuse vs No known issues	1.12	0.11	0.11	1.0	0.32	-0.11	0.33
Other children with care and protection history	Yes vs No	1.74	0.56	0.04	189.1	<.0001	0.48	0.64
Single parent	Yes vs Not single or unknown	1.16	0.15	0.05	9.5	0.00	0.05	0.25
Single parent	Single and no father listed vs not single or unknown	1.17	0.16	0.07	5.1	0.02	0.02	0.30
Parenting demands	High parenting demands vs no high parenting demands	1.15	0.14	0.05	8.1	0.00	0.04	0.23
Parenting demands	No other children vs no parenting demands	1.04	0.04	0.05	0.6	0.44	-0.06	0.14
Family violence	Events in one of the last 12 months vs no events	1.03	0.03	0.09	0.1	0.73	-0.14	0.20
Family violence	Events in more than one of the last 12 months vs no events	1.67	0.51	0.10	25.2	<.0001	0.31	0.71
Rate of investigation	Low vs High	1.01	0.01	0.04	0.0	0.90	-0.08	0.09
Rate of investigation	Medium vs High	0.89	-0.12	0.04	6.8	0.01	-0.20	-0.03

(w) Model with aggregated local level variables based on multiple counts of notifications and investigations

Variable	Categories	Odds Ratio	Estimate	SE	Chi-Square	Sign.	95% CI Lower	95% CI Upper
Benefit address changes in last year	No address changes vs Missing/no benefit in last year	0.82	-0.20	0.07	7.4	0.01	-0.34	-0.05
Benefit address changes in last year	1 or 2 address changes vs Missing/no benefit in last year	0.78	-0.25	0.07	12.4	0.00	-0.39	-0.11
Benefit address changes in last year	3 plus address changes vs Missing/no benefit in last year	1.36	0.31	0.10	10.4	0.00	0.12	0.50
Benefit caregiver is not birth registration parent	Yes vs No	1.98	0.68	0.13	25.8	<.0001	0.43	0.96
Benefit caregiver is not birth registration parent	No birth registration vs No	0.99	-0.01	0.08	0.0	0.91	-0.18	0.15
Caregiver with care and protection history	Yes vs No	1.30	0.26	0.04	38.5	<.0001	0.18	0.35
Correction history in the last 5 years	Non-custodial sentence vs No history	1.09	0.09	0.08	1.2	0.27	-0.08	0.24
Correction history in the last 5 years	Custodial Sentence for non-violent crimes vs No history	0.90	-0.10	0.12	0.8	0.38	-0.33	0.13
Correction history in the last 5 years	Custodial Sentence for violent crimes vs No history	1.42	0.35	0.10	11.4	0.00	0.16	0.57
Caregiver's age	Under 20 vs 30_34/missing	1.48	0.39	0.09	21.2	<.0001	0.23	0.56
Caregiver's age	20-24 vs 30-34	0.86	-0.15	0.07	5.3	0.02	-0.27	-0.01
Caregiver's age	25-29 vs 30-34	0.87	-0.14	0.07	3.8	0.05	-0.26	0.02
Caregiver's age	35-39 vs 30-34	0.97	-0.03	0.10	0.1	0.77	-0.23	0.18
Caregiver's age	40 plus vs 30-34	1.22	0.20	0.14	2.1	0.15	-0.10	0.45
Behavioural or relationship difficulties as a child	Yes vs No	1.22	0.20	0.05	18.7	<.0001	0.11	0.29
Time on benefit in last 5 years	More than 80% vs no time	2.00	0.69	0.07	104.0	<.0001	0.56	0.83
Time on benefit in last 5 years	20<-80% vs no time	1.83	0.61	0.06	107.4	<.0001	0.49	0.72
Time on benefit in last 5 years	Up to 20% vs no time	1.10	0.10	0.08	1.5	0.21	-0.04	0.26
Mental health in last 5 years	Substance abuse vs No known issues	1.25	0.22	0.12	3.5	0.06	-0.04	0.43
Mental health in last 5 years	Persistent substance abuse issues vs No known issues	1.59	0.47	0.15	9.5	0.00	0.23	0.83
Mental health in last 5 years	Mental health issues other than substance abuse vs No known issues	0.74	-0.30	0.09	10.8	0.00	-0.49	-0.13
Mental health in last 5 years	Persistent mental health issues other than substance abuse vs No known issues	1.14	0.13	0.11	1.4	0.23	-0.09	0.35
Other children with care and protection history	Yes vs No	1.74	0.56	0.04	190.0	<.0001	0.48	0.64
Single parent	Yes vs Not single or unknown	1.15	0.14	0.05	8.6	0.00	0.05	0.24
Single parent	Single and no father listed vs not single or unknown	1.18	0.16	0.07	5.5	0.02	0.03	0.31
Parenting demands	High parenting demands vs no high parenting demands	1.15	0.14	0.05	8.3	0.00	0.05	0.23
Parenting demands	No other children vs no parenting demands	1.03	0.03	0.05	0.4	0.52	-0.07	0.13
Family violence	Events in one of the last 12 months vs no events	1.01	0.01	0.09	0.0	0.90	-0.17	0.17
Family violence	Events in more than one of the last 12 months vs no events	1.70	0.53	0.10	27.2	<.0001	0.34	0.74
Rate of investigation	Low vs High	0.96	-0.04	0.05	0.7	0.40	0.01	0.21
Rate of investigation	Medium vs High	0.88	-0.13	0.04	8.5	0.00	-0.25	-0.07

Appendix 5 – Profile of 3000 children with highest PRM scores of birth cohort 2007 using the algorithm developed on cohort 2010

		Profile of all children of cohort 2007	Profile of children of cohort 2007 with incidence of maltreatment by age 2
Numbers to be treated		62,273	1,434
Gender of child	Male	51.5	53.0
	Female	48.5	47.0
Low birth weight or pre-term	Yes	8.2	10.7
	No or unknown	91.8	89.3
Parenting demands	High parenting demands	20.1	32.9
	No other children	59.4	51.1
	Other children but not high parenting demands	20.5	16.0
Other children with care and protection history	Yes	4.8	34.9
	No	95.3	65.1
Family violence	Events in one month	0.6	5.6
	Events in more than one month	0.2	3.0
	No events (no Police FV notifications or contact records)	99.3	91.4
Caregiver's age	Under 20	7.8	22.4
	20 to 25	17.5	31.1
	25 to 30	24.0	20.4
	30 to 35(includes missing)	28.6	14.0
	35 to 40	18.1	8.5
	Over 40	4.0	3.6
Benefit caregiver is not a birth registration parent	Yes	0.8	2.3
	No birth registration	6.3	26.1
	No	92.9	71.6
Single parent	Single parent	19.4	59.3
	Single parent and no father listed on birth registration	5.2	15.1
	Not single parent or partnership status unknown	75.4	25.7
Time on benefit	More than 80%	11.7	50.7
	20<-80%	17.0	31.0
	Up to 20%	14.1	11.1
	No time	57.1	7.3
Caregivers with care and protection history	Yes	9.8	43.2
	No	90.2	56.8
Benefit caregiver's address changes in the last year	No address changes	23.0	17.7
	1-2addresschanges	9.8	16.2
	3plusaddresschanges	1.8	9.9
	Missing(no benefit in last year)	65.4	56.2
Mental health	Substance abuse issues	1.1	6.0
	Persistent substance abuse issues(3+yearsinlast5)	0.4	2.2
	Mental health issues other than substance abuse	3.6	10.5
	Persistent mental health issues other than substance abuse (3+years in last 5)	1.3	6.9
	No known mental health or substance abuse issues	93.6	74.4
Behavioural or relationship difficulties as a child	Yes	3.5	19.1
	No	96.5	80.9
Caregivers' Corrections history in the last 5 years	Non-custodial sentence	4.1	14.4
	Custodial sentence for non-violent crimes	1.5	4.3
	Custodial sentence for violent crimes	1.7	6.8
	No history	92.8	74.5

(a) profile of 3000 children with highest PRM score based on base models

		Base model for cohort 2008 applied on cohort 2007	Base model for cohort 2009 applied on cohort 2007	Base model for cohort 2010 applied on cohort 2007
Numbers to be treated		3,000	3,000	3,000
Gender of child	Male	52.2	52.6	52.5
	Female	47.8	47.4	47.5
Low birth weight or pre-term	Yes	8.9	8.4	8.2
	No or unknown	91.1	91.6	91.8
Parenting demands	High parenting demands	41.9	33.5	38.1
	No other children	36.4	56.3	45.7
	Other children but not high parenting demands	21.7	10.2	16.2
Other children with care and protection history	Yes	78.6	53.3	59.9
	No	21.4	46.7	40.1
Family violence	Events in one month	9.8	5.9	8.8
	Events in more than one month	3.9	3.7	3.7
	No events (no Police FV notifications or contact records)	86.3	90.4	87.6
Caregiver's age	Under 20	11.3	32.9	27.0
	20 to 25	32.6	26.8	27.6
	25 to 30	25.8	21.2	21.4
	30 to 35(includes missing)	15.5	10.0	9.9
	35 to 40	10.2	6.1	8.6
	Over 40	4.5	3.0	5.4
Benefit caregiver is not a birth registration parent	Yes	2.4	3.5	6.2
	No birth registration	30.9	29.6	35.5
	No	66.7	66.9	58.4
Single parent	Single parent	73.7	74.4	72.8
	Single parent and no father listed on birth registration	14.8	12.5	14.9
	Not single parent or partnership status unknown	11.5	13.0	12.3
Time on benefit	More than 80%	76.6	62.5	65.5
	20<-80%	21.8	35.5	31.7
	Up to 20%	1.6	2.0	2.7
	No time	0.0	0.0	-
Caregivers with care and protection history	Yes	52.1	63.0	57.2
	No	47.9	37.0	42.8
Benefit caregiver's address changes in the last year	No address changes	18.7	10.0	17.5
	1-2addresschanges	17.2	17.7	15.0
	3plusaddresschanges	9.2	16.0	15.4
	Missing(no benefit in last year)	54.9	56.3	52.1
Mental health	Substance abuse issues	10.2	10.2	9.4
	Persistent substance abuse issues(3+yearsinlast5)	4.2	5.0	4.7
	Mental health issues other than substance abuse	13.2	16.6	10.7
	Persistent mental health issues other than substance abuse (3+years in last 5)	9.7	8.3	8.2
	No known mental health or substance abuse issues	62.7	59.9	66.9
Behavioural or relationship difficulties as a child	Yes	22.2	26.8	29.1
	No	77.8	73.2	70.9
Caregivers' Corrections history in the last 5 years	Non-custodial sentence	21.3	19.5	17.0
	Custodial sentence for non-violent crimes	8.3	8.8	6.6
	Custodial sentence for violent crimes	5.3	9.7	10.0
	No history	65.1	62.1	66.4

(b) profile of 3000 children with highest PRM score from models testing sensitivity to various administrative data and linkages

		Model based on benefit and care protection data only	Model based on benefit, care and protection data and Corrections data	Model for cohort 2010 using less conservative link 3	Model based on benefit, care and protection, Corrections and Health data
Numbers to be treated		3,000	3,000	3,000	3,000
Gender of child	Male	51.8	51.6	53.0	51.8
	Female	48.2	48.4	47.0	48.2
Low birth weight or pre-term	Yes	-	-	7.9	14.6
	No or unknown	-	-	92.1	85.4
Parenting demands	High parenting demands	37.0	39.1	37.3	45.4
	No other children	43.0	40.1	41.8	40.3
	Other children but not high parenting demands	20.0	20.8	20.9	14.2
Other children with care and protection history	Yes	57.3	56.2	68.5	58.2
	No	42.7	43.8	31.5	41.8
Family violence	Events in one month	7.9	7.7	8.8	5.9
	Events in more than one month	3.3	3.1	4.0	5.8
	No events (no Police FV notifications or contact records)	88.8	89.1	87.2	88.3
Caregiver's age	Under 20	30.6	27.8	28.9	20.2
	20 to 25	30.4	33.4	31.6	37.1
	25 to 30	17.2	17.9	17.7	22.3
	30 to 35(includes missing)	9.9	9.8	11.0	10.4
	35 to 40	7.4	7.9	6.0	7.4
	Over 40	4.5	3.2	4.8	2.6
Benefit caregiver is not a birth registration parent	Yes			1.7	2.1
	No birth registration	na		37.9	29.7
	No			60.5	68.2
Single parent	Single parent	86.0	89.5	71.6	63.8
	Single parent and no father listed on birth registration	-	-	20.4	14.7
	Not single parent or partnership status unknown	14.0	10.5	7.9	21.4
Time on benefit	More than 80%	62.6	62.9	76.1	61.6
	20<-80%	34.7	35.3	23.1	36.9
	Up to 20%	2.6	1.6	0.8	1.5
	No time	0.1	0.1	-	-
Caregivers with care and protection history	Yes	53.7	54.2	64.2	69.8
	No	46.3	45.8	35.8	30.2
Address changes in the last year	No address changes	28.7	26.9	19.8	29.9
	1-2 address changes	38.3	40.1	19.4	49.7
	3 plus address changes	33.0	32.9	10.7	17.3
	Missing	-	0.0	50.0	3.1
Mental health	Substance abuse issues	5.9	6.3	9.7	21.5
	Persistent substance abuse issues(3+years in last 5)	2.4	1.9	4.0	6.6
	Mental health issues other than substance abuse	9.8	9.7	10.6	21.7
	Persistent mental health issues other than substance abuse (3+years in last 5)	7.0	6.9	7.3	10.0
	No known mental health or substance abuse issues	74.9	75.2	68.4	40.3
Behavioural or relationship difficulties as a child	Yes	19.8	21.5	26.0	29.3
	No	80.2	78.5	74.0	70.7
Caregivers' Corrections history in the last 5 years	Non-custodial sentence	-	9.6	23.2	13.2
	Custodial sentence for non-violent crimes	-	3.6	9.8	5.4
	Custodial sentence for violent crimes	-	4.6	9.6	6.7
	No history	-	82.2	57.4	74.7

(c) profile of 3000 children with highest PRM score from models to correct for disproportionality

		Model for Māori	Model for "Non-Māori" and ethnic group not recorded	Model where no other children with history of contact with CYF	Model where other children with history of contact with CYF
Numbers to be treated		1,770	1,230	1,950	1,050
Gender of child	Male	54.0	50.7	51.6	52.5
	Female	46.1	49.4	48.4	47.5
Low birth weight or pre-term	Yes	7.9	8.5	16.8	6.3
	No or unknown	92.1	91.5	83.2	93.7
Parenting demands	High parenting demands	36.8	25.5	19.6	47.0
	No other children	48.0	59.4	73.6	23.5
	Other children but not high parenting demands	15.2	15.2	6.8	29.5
Other children with care and protection history	Yes	60.9	42.5	-	100.0
	No	39.1	57.5	100.0	-
Family violence	Events in one month	9.2	8.1	1.7	15.9
	Events in more than one month	3.3	4.2	0.2	10.2
	No events (no Police FV notifications or contact records)	87.5	87.7	98.2	73.9
Caregiver's age	Under 20	31.5	38.5	43.8	8.7
	20 to 25	27.3	30.1	34.9	37.6
	25 to 30	21.1	10.7	12.1	31.1
	30 to 35 (includes missing)	9.4	8.8	4.1	12.5
	35 to 40	6.3	7.1	3.2	6.7
Benefit caregiver is not a birth registration parent	Yes	5.0	1.5	3.7	2.1
	No birth registration	38.1	25.9	17.2	51.7
Single parent	No	56.9	72.6	79.1	46.2
	Single parent	75.3	62.4	73.1	72.3
	Single parent and no father listed on birth registration	14.7	20.7	14.0	13.8
Time on benefit	Not single parent or partnership status unknown	10.0	17.0	12.9	13.9
	More than 80%	68.3	50.9	54.1	88.4
	20<-80%	29.8	38.7	40.0	11.1
	Up to 20%	1.8	10.1	5.9	0.5
Caregivers with care and protection history	No time	0.1	0.3	0.1	-
	Yes	59.4	60.0	68.6	72.4
Benefit caregiver's address changes in the last year	No	40.6	40.0	31.4	27.6
	No address changes	16.5	17.1	7.8	22.2
	1-2 address changes	20.1	15.5	8.6	23.5
	3 plus address changes	12.0	9.3	11.3	14.0
Caregivers' mental health issues	Missing (no benefit in last year)	51.5	58.2	72.4	40.3
	Substance abuse issues	6.5	12.1	11.7	6.5
	Persistent substance abuse issues (3+years in last 5)	2.9	7.2	4.1	2.4
	Mental health issues other than substance abuse	11.4	12.4	20.5	10.3
	Persistent mental health issues other than substance abuse (3+years in last 5)	7.9	7.0	9.6	7.2
Behavioural or relationship difficulties as a child	No known mental health or substance abuse issues	71.3	61.3	54.1	73.6
	Yes	23.9	33.0	34.2	26.0
Caregivers' Corrections history in the last 5 years	No	76.1	67.0	65.8	74.0
	Non-custodial sentence	17.8	14.0	16.5	12.9
	Custodial sentence for non-violent crimes	8.3	5.0	12.5	6.0
	Custodial sentence for violent crimes	8.4	10.3	14.2	5.8
	No history	65.6	70.7	56.9	75.3

(d) profile of 3000 children with highest PRM score from models to correct for disproportionality-continued

		Model for Beneficiaries using link 3c	Model for Beneficiaries using link 4c	Model for non-Beneficiaries using link 3c	Model for non-Beneficiaries using link 4c
Numbers to be treated		2,370	2,370	630	630
Gender of child	Male	52.2	53.2	52.2	51.6
	Female	47.8	46.8	47.8	48.4
Low birth weight or pre-term	Yes	10.8	18.3	12.5	13.3
	No or unknown	89.2	81.7	87.5	86.7
Parenting demands	High parenting demands	35.2	50.7	32.4	43.5
	No other children	52.5	39.1	45.6	41.9
	Other children but not high parenting demands	12.3	10.2	22.1	14.6
Other children with care and protection history	Yes	60.1	59.2	29.4	24.0
	No	40.0	40.8	70.6	76.0
Family violence	Events in one month	9.8	8.7	2.9	2.1
	Events in more than one month	4.0	8.9	1.6	2.7
	No events (no Police FV notifications or contact records)	86.2	82.5	95.6	95.2
Caregiver's age	Under 20	32.3	29.2	14.1	29.8
	20 to 25	30.8	32.9	35.1	36.2
	25 to 30	15.6	13.9	27.6	22.1
	30 to 35(includes missing)	8.8	11.5	12.4	7.3
	35 to 40	7.9	9.0	8.9	3.8
	Over 40	4.6	3.5	1.9	0.8
Benefit caregiver is not a birth registration parent	Yes	4.5	4.1	-	-
	No birth registration	29.7	36.2	-	-
	No	65.8	59.7	100.0	100.0
Single parent	Single parent	71.8	72.0	23.8	30.8
	Single parent and no father listed on birth registration	14.0	14.4	7.9	13.3
	Not single parent or partnership status unknown	14.2	13.6	68.3	55.9
Time on benefit	More than 80%	66.2	63.0	49.7	83.0
	20<-80%	31.5	34.1	39.5	16.2
	Up to 20%	2.2	2.9	8.3	0.8
	No time	0.0	0.1	2.5	-
Caregivers with care and protection history	Yes	67.8	62.6	61.1	92.1
	No	32.2	37.4	38.9	7.9
Address changes in the last year	No address changes	10.8	31.5	8.7	48.1
	1-2 address changes	13.5	46.2	7.0	33.5
	3 plus address changes	15.3	16.3	1.4	1.9
	Missing	60.5	6.0	82.9	16.5
Caregivers' mental health issues	Substance abuse issues	10.0	23.2	9.7	7.0
	Persistent substance abuse issues(3+years in last 5)	3.9	6.1	6.4	1.4
	Mental health issues other than substance abuse	11.9	21.1	17.0	19.8
	Persistent mental health issues other than substance abuse (3+years in last 5)	9.4	10.5	10.0	7.5
	No known mental health or substance abuse issues	64.9	39.2	57.0	64.3
Behavioural or relationship difficulties as a child	Yes	36.6	33.0	26.0	35.7
	No	63.4	67.0	74.0	64.3
Caregivers' Corrections history in the last 5 years	Non-custodial sentence	19.0	18.7	14.0	12.7
	Custodial sentence for non-violent crimes	9.8	8.6	33.3	6.4
	Custodial sentence for violent crimes	7.8	8.1	14.0	5.2
	No history	63.5	64.7	38.7	75.7

(e) profile of 3000 children with highest PRM score from models targeting different outcomes

		Model to predict substantiated physical or sexual abuse or neglect (excluding emotional abuse findings)	Model to predict investigations or Child and Family Assessments	Model to predict substantiated physical abuse or maltreatment or marker injury hospitalisation	Model to predict notifications (including Police FV notifications and contact records)
Numbers to be treated		3,000	3,000	3,000	3,000
Gender of child	Male	51.8	51.3	51.5	52.1
	Female	48.2	48.7	48.5	47.9
Low birth weight or pre-term	Yes	7.2	9.4	6.9	11.5
	No or unknown	92.8	90.6	93.1	88.5
Parenting demands	High parenting demands	47.3	26.7	37.5	32.4
	No other children	35.4	57.2	39.9	54.2
	Other children but not high parenting demands	17.3	16.0	22.7	13.4
Other children with care and protection history	Yes	86.2	50.6	62.3	55.3
	No	13.8	49.4	37.7	44.7
Family violence	Events in one month	8.5	5.1	6.3	8.6
	Events in more than one month	3.2	2.1	2.6	3.6
	No events (no Police FV notifications or contact records)	88.2	92.8	91.1	87.8
Caregiver's age	Under 20	12.7	44.5	33.3	34.2
	20 to 25	29.0	29.8	24.7	32.1
	25 to 30	24.7	12.0	19.0	16.1
	30 to 35(includes missing)	16.6	5.6	8.5	8.0
	35 to 40	11.7	3.3	11.7	6.3
Benefit caregiver is not a birth registration parent	Yes	4.2	2.1	9.4	2.8
	No birth registration	44.8	25.0	45.7	28.1
Single parent	No	51.1	73.0	44.8	69.2
	Single parent	67.6	76.1	68.3	76.2
	Single parent and no father listed on birth registration	15.2	17.3	14.6	13.9
Time on benefit	Not single parent or partnership status unknown	17.3	6.6	17.1	9.9
	100-80%	69.1	60.7	76.4	60.2
	20-80%	30.4	37.4	20.4	36.3
	Under20%	0.5	1.9	3.1	3.5
Caregivers with care and protection history	No time	-	-	0.1	-
	Yes	46.3	78.0	65.7	70.9
Address changes in the last year	No	53.7	22.0	34.3	29.1
	No address changes	24.6	13.0	26.0	9.4
	1-2addresschanges	22.5	15.1	23.6	15.4
	3plusaddresschanges	12.1	9.6	11.8	14.2
Mental health	Missing(no benefit in last year)	40.8	62.3	38.6	60.9
	Substance abuse issues	5.8	10.1	4.2	11.6
	Persistent substance abuse issues (3+years in last 5)	2.1	4.7	1.5	4.6
	Mental health issues other than substance abuse	10.8	16.6	9.3	16.2
	Persistent mental health issues other than substance abuse (3+years in last 5)	6.1	8.0	4.9	8.3
Behavioural or relationship difficulties as a child	No known mental health or substance abuse issues	75.1	60.6	80.2	59.3
	Yes	17.1	32.1	24.4	37.1
Caregivers' Corrections history in the last 5 years	No	82.9	67.9	75.6	62.9
	Non-custodial sentence	11.7	13.3	11.0	20.2
	Custodial sentence for non-violent crimes	5.0	5.4	4.7	10.0
	Custodial sentence for violent crimes	5.7	5.7	5.3	9.9
	No history	77.6	75.5	78.9	59.9

References

Belsley, D. A., Edwin. K., and Welsch, R.E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: John Wiley and Sons.

Chawla, Nitesh V. (2010) [Data Mining for Imbalanced Datasets: An Overview](#) In: Maimon, Oded; Rokach, Lior (Eds) *Data Mining and Knowledge Discovery Handbook*, Springer (pages 875-886)

Choi, J. M., (2010). A selective sampling method for imbalanced data learning on support vector machines. PhD Dissertation. Iowa State University, USA. Available online lib.dr.iastate.edu/cgi/viewcontent.cgi?article=2549&context=etd

Department of Corrections (2012). *Offender Volumes Report 2011*. Wellington: Department of Corrections. Available from: http://www.corrections.govt.nz/research/offender-volumes-report/offender_volumes_report_2011.html

Estabrooks, A., Jo, T. & Japkowicz, N. (2004). A Multiple Resampling Method for Learning from Imbalanced Data Sets. *Computational Intelligence*. Vol. 20, pp 18-36.

Ferron, J. et. al. (2004). Hierarchical Linear Modeling: A Review of Methodological Issues and Applications. Annual meeting of the American Educational Research Association, Sand Diego, April 12-16, 2004.

Japkowicz, N.(2001). Supervised versus unsupervised binary-learning by feed forward neural networks, *Machine Learning* **42** (1–2), 97–122

Laurikkala, J. (2001). Improving identification of difficult small classes by balancing class distribution. *Proceedings of AI in Medicine in Europe: Artificial Intelligence Medicine*, pp. 63-66.

Ministry of Social Development (2012). *Statistical Report for the Year Ending 2011*. Wellington: Ministry of Social Development. Available from: <http://www.msd.govt.nz/about-msd-and-our-work/publications-resources/statistics/statistical-report/statistical-report-2011.html>

Montgomery, D. C., Peck, E. A., and Vining, G. G. (2001). *Introduction to linear regression analysis*, 3rd edition, Wiley, New York.

Putnam-Hornstein, E., and Needell, B. (2011). Predictors of child welfare contact between birth and age five: an examination of California's 2002 birth cohort. *Children and Youth Services Review*, 33(11), 2400-2407.

Salmond, C., Crampton, P., & Atkinson, J. (2007). *NZDep2006 Index of Deprivation User's Manual*. Wellington: Department of Public Health, University of Otago.

Statistics New Zealand (2010). *Late Birth Registrations*. Wellington: Statistics New Zealand. Available from: http://www.stats.govt.nz/browse_for_stats/population/births/late-birth-registrations.aspx

Weiss, G. (2004). Mining with rarity: A unifying framework. *SIGKDD Explorations* 6 (1). pp 7-19.

Witten, Frank E. and Hall M., 2011: *Data mining: practical machine learning tools and techniques* 3rd edition 2011. Morgan Kaufman Publishers.

Williams, 2011: *Data mining with Rattle and R: The art of excavating data for knowledge discovery*. Springer

Wu, S., Ma, C., Carter, R., Ariet, M., Feaver, E., Resnick, M., and Roth, J. (2004). Risk factors for infant maltreatment: a population-based study. *Child Abuse & Neglect*, 28, 1253–1264.