NZ crime & safety SURVEY

2014 Technical Manual

Published by the Ministry of Justice 2015 © Crown Copyright
ISBN 978-0-478-32452-5 (print)
ISBN 978-0-478-32453-2 (online)

Ministry of Justice
The Vogel Centre, 19 Aitken Street
DX SX10088, Wellington, New Zealand

T +64 4 918 8800
F +64 4 918 8820
E research@justice.govt.nz
W www.justice.govt.nz/NZ-CRIME-AND-SAFETY-SURVEY

Contents

Tables

Figures

Executive summary

The following table provides a summary of the key methodological elements of the New Zealand Crime and Safety Survey (NZCASS) 2014.

	Details	
Overview	Nationwide, face-to-face random probability survey, with 1 respondent selected per household using multistage stratified cluster sampling methods.	
Target population	Total usually resident, non-institutionalised, civilian population of New Zealand aged 15 years and over.	
Sampled areas	North Island, South Island and Waiheke Island.	
Dwellings included	Permanent, private dwellings. **Note:** While hospitalised or dependent residents of homes for the elderly were ineligible for the survey (ie living in institutions), residents of aged care facilities who were living independently in a permanent, private dwelling (eg a self-contained unit) were eligible.	
Sample composition	Two samples are drawn as part of the NZCASS: a general or 'main sample' and a Māori booster sample that aims to increase sample sizes for Māori.	
Sample size	Main sample:	5,235
	Māori booster sample:	1,708
	Total sample:	6,943
Response rates	Main sample:	80.0%
	Māori booster sample:	84.8%
	Total sample:	81.0%
Interviewing period	10 February 2014 – 6 July 2014	
Average interview length	Total questionnaire	40 minutes and 45 seconds
Questionnaire recall period	1 January 2013 to date of interview[1]	
Crimes/offences	In the NZCASS, questions are asked about different things (incidents) that might have happened to the respondent or their household. These incidents are then coded by legal experts to determine whether or not the incident was a crime, and what type of offence (or offences) occurred. Important: **The NZCASS does *not* ask survey participants about crimes that happened to them**. This is because people don't always: • view some things that happen as crimes • know what are legally considered crimes and what aren't.	

[1] While most critical questions use the recall period 1 January 2013 to the date of the interview, there are some that refer to a different period (eg the lifetime prevalence questions relating to offences by a partner).

	Details
Comparability between surveys	The 2014 project team took great care to maintain comparability between survey years. The NZCASS methodology and application has remained largely consistent with the 2006 and 2009 iterations of the survey, but some changes have been made in order to: • improve response rates • improve project efficiency and transparency • bring questions and analysis in line with Statistics NZ standards and classifications • better meet users' information needs.
Weighting	Three types of weighting are applied: • household and individual weights: to ensure results represent the New Zealand population • incident weights: to adjust for the fact that detailed information was only collected for up to 6 incidents ('victim form information').
Imputation	Data underwent a complex imputation process as detailed information was not collected about all incidents reported in the questionnaire[2].

[2] Detailed information about all offences was not collected in order to reduce respondent burden for highly victimised people.

1. Introduction

The New Zealand Crime and Safety Survey (NZCASS) provides information for researchers, policy makers and the public about the nature and extent of crime and victimisation in New Zealand.

2014 is the third time that the NZCASS has been conducted in its current form, with previous iterations in 2006 and 2009.

The purpose of this manual is to provide:

- a detailed description of the design and methods used
- information about the management and quality assurance processes undertaken as part of the 2014 NZCASS
- Additional technical and analytical information for use of NZCASS findings.

Research objectives

The research objectives of the 2014 NZCASS are to:

- provide information about the extent and nature of crime and victimisation in New Zealand
- measure the extent of crime that goes unreported to Police
- understand who experiences crime and how they respond
- identify the groups at above-average risk of victimisation
- facilitate a better understanding of victims' experiences and needs
- provide a measure of crime trends in New Zealand.

The NZCASS process

The 2014 NZCASS was managed in line with the Projects in a Controlled Environment (PRINCE2) project management methods. To improve efficiency and due to time constraints at different points in the process, a number of project stages were run concurrently or overlapped during the course of the research. As such, the high-level timeline shown in Table 1.1 is not linear in nature.

Table 1.1: The NZCASS process - stage timeline

Dates	Project activities	Description
January – March 2013	Start-up and initiation	Project planning and set-up of initiation and governance structures.
April – July 2013	Tendering and contracting	Open tender for all core services: fieldwork, offence coding and statistical services.
April – May 2013	Needs assessment	Stakeholder engagement/needs assessment to determine information needs for 2014 questionnaire development and analysis and report planning.
June – August 2013	Review and development of questionnaire	Review of questionnaire changes recommended at the end of 2009, assessment of questionnaire in light of stakeholder needs and development of questionnaire for testing as part of pilot study.

Dates	Project activities	Description
August – September 2013	Review of sampling and weighting	Assessment of proposed changes to sampling process by project team and technical advisory group.
August 2013	Cognitive testing	Testing of new questions and selected existing questions.
August – November 2013	Questionnaire set-up and testing	The development of the computer-assisted personal interviewing (CAPI)/computer-assisted self-interviewing (CASI) questionnaire in an online testing environment started during the questionnaire development stage.
November – December 2013	Pilot study	All research related systems and processes tested including but not limited to: fieldwork/interviewing; offence coding; coding quality assurance; data compiling, checking and processing activities.
January 2014	Preparation for main study	Updates to questionnaire, process and systems based on pilot study findings.
February – July 2014	Main study fieldwork	Two interim datasets were provided during the main study: • n200 • n1,000
February – July 2014	Offence coding	Due to the introduction of new technology and processes, offence coding took place at the same time as fieldwork in 2014.
July – September 2014	Data processing	Data cleaning, compiling and formatting datasets, data quality assurance processes.
August – December[3] 2014	Weighting and imputation	• weighting, imputation and quality assurance for 2014 methodology, code and associated deliverables • updated imputation and quality assurance for 2006 and 2009 code and associated deliverables.
October 2014 – June 2015	Analysis and report writing	Analysis and reporting activities started based on 2009 and n1,000 datasets. Continued and finalised once complete 2014 data was delivered.
July – October 2015	Release processes	Sector and public release of core reporting products.
October – December 2015	Project close processes	Benefits realisation, project review and close documentation produced for the ministry.

[3] Imputation errors were discovered in February 2015 during the analysis and reporting stage. This prompted a second round of imputation review and quality assurance, which was completed in April 2015.

Quality assurance processes

Due to the complexity of the NZCASS, specialised quality assurance processes were designed for each different activity and put in place at each stage of the project. These processes have been detailed within each chapter where relevant.

Comparability between surveys

The ability to assess trends and hence comparability between surveys is a key requirement of NZCASS. It is also important to improve the research tools and processes where possible, and ensure that the results produced are useful and relevant to users. Balancing these sometimes-competing requirements is often a challenge.

The project team in 2014 took great care to assess each potential change in relation to its impact on comparability. Where changes have been made, these were carefully evaluated by the project team and experts (where required), and then approved by the NZCASS Steering Group.

Where changes affecting the analysis of information were made in 2014 (eg changes to weighting and imputation), these changes were retrospectively applied to 2006 and 2009 datasets and/or processes to ensure that comparisons between years were possible.

Table 1.2 provides a high-level overview of the main elements of this report along with a note on whether a change has occurred and where you can find more information about that change.

Table 1.2: Overview of changes made for the 2014 NZCASS

Element	Change	Page reference
Sampling	Yes: Introduction of the use of the NZ Post 'Postal Address File' (PAF) and Māori electoral roll as part of the sampling process.	Page 19-24
Questionnaire	Yes: Some changes were made to the questionnaire to bring questions in line with stakeholder needs and (where possible) align demographics with Statistics NZ standard classifications.	Page 33-35
Incident selection	No	N/A
Fieldwork processes	Yes: Use of updated technology to manage fieldwork processes.	Page 47
Fieldwork statistics	Response rates have increased over time both between 2006 and 2009, and between 2009 and 2014.	Page 67
Offence coding	No changes to coding method or rules used but the introduction of updated technology enabled new coding management and quality assurance processes to be implemented. The Offence Coding Manual was also updated to provide clearer and more easily understandable instructions to coders.	Page 80
Data processing	Yes: Due to the change in survey software used, new data processing steps were implemented to ensure data was clean and correct.	Page 94
Classifications, groupings and standards	Yes: All classifications and standards were reviewed as part of the 2014 NZCASS and where possible brought in line with Statistics NZ standards/classifications or other model government research.	Page 97

Element	Change	Page reference
Weighting	Yes: New process added to weighting procedure (linked to changes to the sampling process).	Page 109
Variance estimation	No	N/A
Imputation	Yes: The number of values imputed for each missing value was increased from 10 to 100. This was done to increase the power of significance tests. A number of corrections to the imputation code were made and around 40 new imputation items were added to enable analysis by the victim's relationship to the offender for violent interpersonal offences.	Page 122-123, 136
Reporting analysis	Yes: Adjustments made to Police offence statistics to help improve comparability with NZCASS estimates.	Page 164

2. Sampling

Overview

Table 2.1 provides an overview of the key information relating to the sampling process for the 2014 NZCASS.

Table 2.1: Overview of sampling information

	2014
Overview	Nationwide, face-to-face random probability survey, with 1 respondent selected per household using multistage cluster sampling methods.
Multistage sampling	The sample was drawn in a multistage sequence that started from an unstratified area frame, through to clusters of dwellings, then to a single respondent from each dwelling, and finally a limited number of incidents from each respondent: 1. **selected first:** areas (meshblocks) 2. **selected second:** households (dwellings) within areas 3. **selected third:** one respondent within each household 4. **final:** selection of a small number of incidents from those experienced by respondents.
Samples	Two samples were drawn in 2014: • main sample • Māori booster sample. The purpose of the Māori booster sample was to ensure that the survey produced more reliable results for Māori.
Primary sampling unit (PSU)	Statistics NZ meshblocks.[4]
Number of PSUs selected	One thousand meshblocks were selected with probability proportional to size.
Target population	Total usually resident[5], non-institutionalised, civilian population of New Zealand aged 15 years and over.
Sampled areas	North Island, South Island and Waiheke Island.
Areas excluded	• offshore islands other than Waiheke Island • PSUs containing fewer than 9 dwellings.
Dwellings included	Permanent, private dwellings. **Note:** While hospitalised or dependent residents of homes for the elderly were ineligible for the survey (ie living in institutions), residents of aged care facilities who were living independently in a permanent, private dwelling (eg a self-contained unit) were eligible.
Dwellings excluded	• temporary private dwellings • non-private dwellings. The Māori booster sample only includes addresses where an elector of Māori descent resides.

[4] 2013 meshblock definitions used for main study sampling.
[5] The Australian Bureau of Statistics methodological review of the 1996 National Survey of Crime Victims recommended that the official definition of 'usually resident' be adopted. This has been used for NZCASS and its predecessors consistently since 2001.

	2014
Eligible respondents	As noted above under 'Target population', eligible respondents were usually resident, non-institutionalised civilians, aged 15 years and over. For the Māori booster sample, the respondent also had to identify as Māori to be eligible.
Ineligible respondents	• those who were present at the time of the interview but usually resided elsewhere (either within New Zealand or overseas) • non-New Zealand diplomats and their non-New Zealand staff • members of the non-New Zealand armed forces stationed in New Zealand • overseas visitors in New Zealand for less than 12 months • children under shared custody arrangements if they spent more nights of the week elsewhere. • those living in institutions, hospitals, barracks etc • those without a usual residence (homeless)
Sampling error	Sampling error arises because only a small part of the New Zealand population is surveyed, rather than the entire New Zealand population (census). Because of this, the results (estimates) of the survey will generally differ to some extent from the figures for the entire New Zealand population. This difference due to random sampling variation is known as sampling error. The size of the sampling error depends on the sample size, the size and nature of the estimate, and the design of the survey.

Changes to sampling

Historical overview

The Table 2.2 provides an overview of the main changes to NZCASS sampling since 2006.

Table 2.2: Changes to NZCASS sampling since 2006

Year	Main change(s) from previous iteration	Rationale
2009	PSUs changed from Nielsen Area Units (NAUs) to Statistics NZ meshblocks.	NAUs were the PSUs used by AC Nielsen ('Nielsen') in 2006. In 2009, a new fieldwork provider was contracted to undertake the NZCASS and with this, a new PSU was instated.
	An unstratified sample of PSUs was selected in 2009, in contrast to a stratified sample in 2006.	The stratification framework used in 2006 was not customised for the NZCASS, but was a standard framework used by Nielsen. It was not believed to increase sampling efficiency greatly, and was dropped in the transition to the new fieldwork provider.
	The number of PSUs selected changed from 800 NAUs selected for the main sample and 320 separate NAUs for the Māori booster sample in 2006 to 1000 meshblocks in total in 2009.	Increasing the number of PSUs for the main sample decreased the clustering effect and thus improved the reliability of NZCASS estimates.
	In 2006, the areas selected for the Māori booster sample were selected independently from the main sample areas. In 2009, the same meshblocks were selected for both the main and the Māori booster samples.[6]	Conducting booster sample interviews within (some of) the PSUs selected for the main sample improved the cost- effectiveness of fieldwork, enabling less tightly clustered samples and a larger booster sample.
2014	Introduction of the use of the NZ Post PAF and Māori electoral roll as part of sampling process. Details of these changes are provided below.	• Improve efficiencies in contacting Māori respondents • help to achieve a higher response rate • reduce fieldwork costs.

2014 changes to sampling

In 2014, the ministry's contracted fieldwork provider, CBG Public Sector Surveying (CBG)[7], proposed the following changes to the NZCASS sampling approach:

- sampling addresses sourced from NZ Post's PAF
- sampling Māori booster dwellings from the Māori descent indicator on the electoral roll[8]
- within the booster sample, no ethnicity screening takes place at the door. This means that respondents are selected based on standard criteria (eg 'person 15 or over with the next birthday in the household') and both Māori and non-Māori interviews take place. The rationale for this proposal was to eliminate any disappointment amongst those living at booster addresses who were not eligible to take part and to create a unified selection process for interviewers to follow.

[6] This change in the design for the Māori booster sample may have had some effects on survey estimates, although these effects are considered to be small. Survey weights were adjusted to accommodate the change. This change is also expected to have increased the Māori booster sample response rate.
[7] CBG Health Research Ltd trading as 'CBG Public Sector Surveying'
[8] This is not the same as choosing to vote in a Māori electorate.

After reviewing the proposed design, James Reilly of Statistical Insights[9] recommended a slightly adapted approach in which screening for ethnicity still took place at the door (as done in the 2009 NZCASS). This meant that there would be no time spent interviewing non-Māori in the Māori booster sample and hence created further cost efficiencies.

Table 2.3 gives a high-level overview of the key differences and similarities between the 2009 approach and the 2014 approach.

Table 2.3: Differences between 2009 and 2014 sampling - main sample and Māori booster

Sample	2009	2014
Main	Dwellings selected by interviewers at fixed intervals along a random route.	Dwellings pre-selected from NZ Post's PAF using an equivalent process to that undertaken in 2009.
Māori booster	Dwellings selected by interviewers at fixed intervals along a random route.	Dwellings pre-selected from addresses on the electoral roll that contain a registered elector of Māori descent.
	The number of dwellings approached remained fixed at 16 across meshblocks.	The number of dwellings approached varied across meshblocks to reflect the proportion of Māori in each meshblock.
	Ask at the door to list all those usually living there aged 15 years or older who might consider themselves Māori.	No change, same approach as in 2009.

2014 change control processes

Overview

All proposed changes were reviewed by technical experts and approved by the NZCASS Steering Group prior to any change being made.

The review process was as follows:

1. proposal made (CBG)
2. proposal reviewed by Statistical Insights in order to assess the impact of the changes and provide a recommendation on whether the changes should be adopted in their original form or with modifications
3. technical advisory group formed to review the proposal and Statistical Insights' assessment, and to provide advice on the change from an expert perspective[10]
4. recommendation made to the NZCASS Steering Group
5. approved by the NZCASS Steering Group.

[9] NZCASS contracted statistical services provider

[10] The technical advisory group consisted of professional statisticians and technical experts from the Ministry of Justice, New Zealand Police, Statistics NZ and the Ministry of Health. James Reilly of Statistical Insights and members of the NZCASS Project Team were also present.

Change risk assessment

The main conclusions reached by the technical advisory group were as follows:

- **Comparability with previous iterations of the survey:** The risks to comparability were sufficiently small that the change would be acceptable if there were sufficient cost savings.
- **Respondent burden:** Moving away from the 2009 approach would reduce respondent burden (ie the number of houses we approach seeking cooperation for a survey but then reject because they have no Māori residents would be reduced).
- **Change risk assessment:** A number of design risks were identified (table 2.4); however, it was agreed that the responses to these risks were sufficient to adequately manage or eliminate the risks effectively.

Table 2.4: Sampling design risks identified

	Design risk	Response	Predicted outcome/effect
1	The PAF is not a complete listing of all permanent private dwellings, especially in rural and highly urbanised areas. This undercoverage means using only the PAF could introduce bias.	CBG will physically enumerate dwellings in a meshblock and add any not covered in the PAF to the sample. Monitor coverage of the PAF.	Bias expected to be reduced to negligible levels.[11]
2	Only 85% of households where Māori live are covered by the electoral roll. Coverage decreases between general elections.	Weighting will ensure that the undercoverage does not result in significant bias. (Māori selected as part of the main sample, not in households identified by the electoral roll, will have larger weights in general than those selected in the booster sample.)	Bias controlled. Weights will become more variable; hence more booster interviews needed than in 2009 to achieve the same margin of error (ie the 2009 approach required a target booster sample of 1,200 whereas the 2014 approach required a target booster sample of 1,660).
3	Not everyone of Māori descent identifies as Māori (ie not everyone selected through the Māori electoral roll will identify as Māori).	Conducting ethnicity screening at the door, and then later in more detail in the questionnaire, will help to ensure that only those who identify as Māori are interviewed as part of the Māori booster.	Screening component of approach is consistent with 2009.
4	Māori aged 15–17 are not covered by the electoral roll.	Those aged 15–17 are likely to be living with someone older who is enrolled.	Bias thought to be negligible.[12]

Overall, it was estimated that the new 2014 approach would increase the design effect for results among Māori by around 25% (ie the margin of error for the same number of Māori interviews would be distinctly higher). In order to compensate for the estimated design effect, a larger sample size (for the Māori booster) is required to ensure that the margin of error remains the same as in 2009:

- **2009 approach:** targeted number of booster interviews = 1,200
- **2014 approach:** targeted number of booster interviews = 1,660.

[11] This risk of bias is accepted in the New Zealand Health Survey, which uses the same sampling approach. CBG estimated that 2–3% of meshblocks would require complete enumeration, and that in other meshblocks only 2–3% of addresses would be enumerated.

[12] It was noted by the Ministry of Health that this bias was within acceptable levels for the New Zealand Health Survey.

Change impact review

The impact of the changes to the sampling process in 2014 can be measured as follows:

- fieldwork costs notably reduced
- increased Māori booster response rate from 69% in 2009 to 85% in 2014[13]
- all design risks monitored and controlled for successfully.

Design risk 1: Undercoverage of PAF increases bias

The risk of undercoverage in the 2014 NZCASS because the PAF is not a complete listing of all occupied private dwellings appeared at first glance to be a minor issue, with only 188 addresses being enumerated.

The number of residential addresses in the selected meshblocks, based on PAF data supplemented by the electoral roll and enumeration, was, however, 2.2% short of the total expected based on 2013 Census counts. A better estimate of the undercoverage, allowing for post-censal population growth, is 3.1%. The calculation of these shortfalls is outlined in Table 2.5. Appendix A gives further details.

It appears enumeration did not address most of the PAF undercoverage and thereby minimise the potential for bias. The potential undercoverage bias should nonetheless be minor, because the PAF had fairly high coverage before enumeration.

Table 2.5: Calculations of PAF shortfalls

		Notes
Number of meshblocks sampled	1,000	
Number of residential addresses in these meshblocks	64,379	From PAF/electoral roll combined
Number of occupied private dwellings in these meshblocks	56,382	2013 Census
Number of addresses selected as part of the main sample, excluding enumerated addresses[14]	7,975	
Number of main sample addresses 'Vacant/Not occupied' or 'Not a dwelling/Empty section', excluding enumerated addresses	1,165	
% of main sample addresses 'Vacant/Not occupied' or 'Not a dwelling/Empty section'	14.6%	
Expected number of residential addresses on PAF/electoral roll in the sampled meshblocks	66,027	Derived by scaling the Census total up to reflect the proportion of addresses that are vacant/unoccupied or non-dwellings
Estimated shortfall, to be handled by enumeration	1,648	

[13] While care was taken in 2014 to use the same fieldwork outcome codes and to calculate the response rate in the same way as in 2009, response rates for different sample designs are not directly comparable.

[14] Enumerated addresses refer to dwellings missing from the PAF list and collected as part of the meshblock survey process.

		Notes
Number of enumerated addresses	188	
Estimated shortfall remaining after enumeration	1,460	
Estimated remaining shortfall as percentage of expected number of residential addresses	2.2%	
Estimated annual growth in number of occupied private dwellings	0.9%	(2006–2013)
Estimated undercoverage	3.1%	From combined PAF/electoral roll database, after enumeration

Design risk 2: 85% of households where Māori live are covered by the electoral roll

Māori not covered by the electoral roll can only enter the main sample, not the booster sample, so they make up a smaller part of the full sample than their share of the population. They have therefore been assigned larger weights than other Māori to ensure they are not under-represented in the survey results.

Māori living in households where no electors of Māori descent reside cannot be surveyed through the booster sample, so we rely on those in the main sample to represent this group. There are proportionately fewer of them in the full sample than there are other Māori, and the survey weights assigned to them need to be larger to compensate and avoid the potential for bias.

The average person weight for Māori selected as part of the main sample, not in households identified by the electoral roll, is 270.8, compared to 181.1 for other Māori. This has ensured that Māori not covered by the electoral roll are not under-represented in the 2014 NZCASS results; their weights account for 18% of the Māori total.

Design risk 3: Not everyone of Māori descent identifies as Māori

Approximately 65% of households contacted in the Māori booster sample were eligible (ie at least 1 person in the household aged 15 or over identified as Māori). In the 2013 Census, however, 84% of people of Māori descent also identified as Māori. The percentage of households containing adults of Māori descent where at least 1 identifies as Māori could well be higher than this, suggesting that a substantial proportion of households containing Māori adults are not identified as such by ethnicity screening at the door.

If these unidentified Māori households tend to differ from those where we proceed to an interview, NZCASS estimates for Māori may be subject to some resultant bias. The doorstep screening process for the Māori booster has not changed since the 2006 NZCASS, so comparisons between the 2006, 2009 and 2014 surveys are unlikely to be greatly affected.

Design risk 4: Māori aged 15–17 are not covered by the electoral roll

As can be seen from Table 2.6, the fourth design risk was not realised. The percentage of booster sample interviews with young Māori respondents (aged 15–17 years) increased from 4.5% in 2009 to 5.1% in 2014.

Table 2.6: Māori sample achieved in main and booster sample – by age

Year	Age	Main sample	Māori booster sample	Total sample
2014	15–17 yrs	28	86	114
	18 yrs or more	647	1,622	2,269
	Age unknown	1	0	1
	Total	**676**	**1,708**	**2,384**
2009	15–17 yrs	34	59	93
	18 yrs or more	509	1,236	1,745
	Age unknown	0	2	2
	Total	**543**	**1,297**	**1,840**
2006	15–17 yrs	35	81	116
	18 yrs or more	474	1,104	1,578
	Age unknown	2	2	4
	Total	**511**	**1,187**	**1,698**

Sampling assumptions, targets and outcomes

This section provides information on:

- the assumptions made in the 2014 NZCASS to design the sample and plan fieldwork
- key targets (eg number of interviews/response rates) and what was achieved
- equivalent statistics for 2006 and 2009 where available.

Assumptions

The assumptions noted in Table 2.7 were used to help estimate statistics like 'the number of interviews expected to be conducted with Māori respondents as part of the main sample' and to help estimate research costs.

Table 2.7: Summary of sampling assumptions, targets and outcomes, by year

	Description	2006	2009	2014
Booster source		Separate multistage area sample	Multistage sample, in areas selected for main sample	Māori electoral roll, in areas selected for main sample
Booster eligibility	Those identifying as Māori who are 15+ years			
Number of PSUs selected	Main sample	800	1,000	1,000
	Māori booster sample	320		
Average interviews per PSU/average interview cluster size	Main sample	5	4	4.8
	Māori booster sample	5	1.41	1.66
Cluster size	Main sample	9	6.5	6.5
Number of trips to each PSU		3	5	5
Maximum number of visits to each dwelling		6+	10	10
Main sample	Non-Māori	3,588	3,458	4,266
	Māori	412	572	543
	Total	**4,000**	**4,030**	**4,809**
Māori booster sample		1,600	1,409	1,660
Estimated average response rate	Main sample	65%	62%	75%
	Māori booster sample	NA	61.5% –65%	75%

NA = Not available

Table 2.8: Summary of interviews targeted and achieved by year

Sample	Respondents	2006		2009		2014	
		Target	Achieved	Target	Achieved	Target	Achieved
Main	Māori respondents[15]	NA	511	572	543	NA	676
	All respondents	4,000	4,229	4,030	4,809	4,809	5,235
	% of respondents identifying as Māori	NA	12.1%	NA	11.3%	NA	12.9%
Māori booster[16]	Māori respondents	1,600	1,187	1,409	1,297	1,660	1,708
Total	Māori respondents	NA	1,698	1,981	1,840	NA	2,384
	All respondents	4,600	5,416	5,439	6,106	6,469	6,943
	% of interviews conducted with Māori	NA	31.4%	NA	30.1%	NA	34.3%

NA = Not available

Booster sample eligibility rates

There were substantial changes in the booster sample's eligibility rate across the 2009 and 2006 iterations of the NZCASS. In 2014, changes to the sample design resulted in a much greater proportion of the booster sample being eligible for selection.

The proportion of occupied dwellings where the initial contact said there was an adult Māori living in their household at the booster eligibility screener question is shown in Table 2.9.

Table 2.9: Booster sample eligibility rates, by year

Year	%	Notes
2006	23%	• In 2006 the screener question changed, incorporating less blunt wording.
2009	17%	• The booster sample design changed in 2009 with booster interviews being conducted in the same areas as the main sample. See the section on 'Changes to sampling' earlier in this chapter. • This change is likely to have caused the eligibility rate to drop. • This change in the sample design has been adjusted for by the survey weights, to enable the comparability of results between the 2006 and 2009 surveys.
2014	61%	• The booster sample design changed in 2014 with Māori booster households being selected from the Māori electoral roll. • 60.8% of eligible respondents within occupied dwellings were selected/contacted. The selected figure captures all outcome codes where eligibility was identified. • The proportion of occupied dwellings was 57.4%, which includes non- contacted addresses where eligibility was not determined.

[15] Māori respondents in the main sample included all respondents who selected Māori as one of their ethnic groups.
[16] Māori respondents interviewed from the Māori booster sample.

Table 2.10: Summary of response rates targeted and achieved, by year

	Sample	2006		2009		2014	
		Target	Achieved	Target	Achieved	Target	Achieved
Response rate	Main	NS	59%	62%	71%	70%	80%
	Māori booster	NS	56%	NS	69%	70%	85%
	Total	**NS**	**58%**[17]	**NS**	**70%**	**70%**	**81%**

NS = Not stated in technical report from year

Survey frame

The survey frame comprises the databases and methods used to select the sample. The first stage in the NZCASS sampling process is to list the meshblocks that fall within the geographical coverage of the survey, and to select a sample of these meshblocks with probability proportional to size. This precedes the selection of dwellings within each meshblock, and respondents within those dwellings.

About meshblocks

A meshblock is the smallest geographical statistical unit for which data is collected and processed by Statistics NZ. Meshblocks provide the aggregation into larger statistical units such as area units, territorial local authorities and regions. The meshblock pattern is reviewed annually. In 2013, there were 46,637 meshblocks defined in New Zealand.

Table 2.11: Inclusions and exclusions from sampling frame

Included	North Island, South Island, Waiheke Island
Excluded	• waterways and inlets • meshblocks with fewer than 9 dwellings.[18]

Meshblocks were selected from both the North and South Islands as well as Waiheke Island. After inclusion and exclusion processes:

- 39,264 meshblocks remained in the frame[19]
- 7373 meshblocks (or 16% of all meshblocks) were excluded
- 1.2% of all occupied private dwellings were excluded from the survey frame.

[17] Not stated in the 2006 Technical Report.
[18] Meshblocks containing fewer than 9 occupied private dwellings were in fact retained in the sampling frame but given zero probability of being selected. This is effectively the same as simply removing them from the frame.
[19] The 2013 Census count of occupied, private dwellings for these meshblocks was 1,542,846.

Primary sampling unit – meshblock selection

The first level of sampling took place at the level of meshblocks. A total of 1,000 meshblocks were selected.

Meshblock selection for the 2014 NZCASS was carried out by Statistics NZ, with guidance from Statistical Insights. Their sampling processes ensure that there is no overlap between the NZCASS sample and areas selected for other samples they manage.

Selection proceeded in 2 steps:

1. 1,000 of the PSUs in Statistics NZ's Household Survey Frame were selected
2. then, 1 meshblock from each PSU was selected.

Both steps involved sampling with probability proportional to the number of occupied private dwellings.[20] As a result, the probability of selection for each meshblock was in direct proportion to the number of occupied private dwellings (as reported in the 2013 Census) within the meshblock. This is known as PPS (probability proportional to size) sampling (Cochran 1977).

The 2009 NZCASS was also based on a systematic PPS sample of meshblocks. In contrast, the 2006 NZCASS was based on separate stratified systematic PPS samples of NAUs, drawn with replacement, for both the main and Māori booster samples.[21]

In 2014, 1,000 meshblocks were chosen, and the main and Māori booster samples were selected from within these meshblocks (see secondary sampling unit). To ensure that interviews were evenly spread over the fieldwork period, approximately 200 meshblocks were scheduled for fieldwork in each of the 5 surveying months (February to June 2014). Meshblocks were initially assigned randomly to months. Some adjustments were made to improve fieldwork efficiency, and the final allocation was profiled to confirm it was evenly spread with respect to crime level, deprivation, region and remoteness.

Secondary sampling unit – dwelling selection

Main sample (core sample)

In each meshblock selected for the 2014 NZCASS, we attempted to select essentially the same number of occupied private dwellings to be approached for the main sample. A systematic sample of dwellings was selected from a list of all dwellings in the meshblock, following the process described in the section titled 'Process for incorporating address files' (page 32). This process distributed the selected dwellings throughout the meshblock.

Part of this process is the selection of every x^{th} address from a randomly selected starting point within the meshblock for the main sample. Here x is the sampling interval, which can be derived by dividing the number of census counts of occupied private dwellings in the meshblock by the cluster size. The cluster size was set at 6.857; that is, the average cluster size of dwellings to be approached in the 1,000 meshblocks for the main sample was 6.857. This cluster size was determined by the number of meshblocks sampled (1,000), the assumed response rate (70%) and the final required

[20] The size of each PSU was set as the sum of the number of occupied private dwellings in its component meshblocks, excluding meshblocks containing fewer than 9 such dwellings. This is equivalent to removing these small meshblocks from the frame and simply using the number of occupied private dwellings as the size measure.

[21] In 2006, a 'with replacement' sampling method was used. One NAU was selected twice for the main sample and another was selected 3 times. No NAUs were selected twice for the Māori booster sample. In 2009 and 2014, there was no practical difference between 'with replacement' and 'without replacement' selection methods, as no meshblocks contained enough dwellings to be selected twice.

sample size (4,800). Approaching 6,857 occupied dwellings with a response rate of 70% would result in 4,800 interviews, so 6.857 dwellings needed to be approached in each meshblock. In all iterations of the NZCASS, the interview cluster size targeted for the main sample has been chosen to provide a good compromise between sample spread and cost efficiency.

As described above, every x^{th} dwelling was included in the main sample, and this method distributed the selected dwellings throughout the meshblock, irrespective of meshblock size. This method is similar to the method used in 2009, and generally produced a smaller clustering effect than in 2006, when every fourth dwelling was approached.[22]

Fieldwork processes

Note regarding changes
In previous iterations of the NZCASS, interviewers were provided with both the description of streets and a map of their assigned meshblock or NAU. Each meshblock or NAU was described according to the streets, side of street and the portion of street belonging to the meshblock or NAU. A systematic, random start point was chosen and marked on each map to prevent interviewers from selecting the start point.

In 2014, dwellings were also selected systematically within each meshblock, but addresses were pre-selected by the fieldwork provider. This meant that interviewers were given a list of addresses they needed to visit prior to arriving at the meshblock. They were also given a complete list of addresses on file for that meshblock, so they could survey the meshblock and enumerate any dwellings that were missing from this list. A proportion of these enumerated dwellings were then selected for the main sample.

Dwellings in rural areas were approached using the same method as that used in non-rural areas. In 2006, consecutive dwellings were approached in rural areas to minimise travel costs. The result of this difference is that there was less interview clustering in rural areas in the 2009 and 2014 surveys than in the 2006 survey.

The final outcome was recorded for every dwelling in the main sample (see Chapter 6 for further details of contact outcomes and response rates in the main sample).[23]

Addresses for the 2014 Māori booster sample were selected from those on the electoral roll where an elector of Māori descent resided, within the 1,000 meshblocks selected for the main sample.

Māori booster sample (screened sample)

2014 NZCASS
Addresses that were already selected for the main sample were excluded. See page 32 for information about the 'Process for incorporating address files'.

Initially the number of booster sample addresses to approach in each meshblock was calculated assuming that 60% of addresses approached would yield a successful interview. As fieldwork progressed, it became clear that this assumption was too optimistic, with the yield for February and March only approaching 50%. It was decided to increase the number of booster sample addresses approached in May and June by 700. This would still achieve the initial target number of booster sample interviews (1,660), assuming the overall yield was around 48%.

[22] As 9 dwellings were approached for each NAU in 2006, this meant that the first 33 dwellings in each NAU were worked in, out of, on average, 230 dwellings (ie around 15% of each NAU was worked in). This differs from the method used in 2009 and 2014, which distributed the selected dwellings across each meshblock.
[23] Outcomes were: Interview (I), Household Refusal (HR), Respondent Refusal (RR), Not Eligible (NE), Access Denied (AD), and Unavailable (U). Please see also Table 6.4 and Table 6.5.

The final outcome was recorded for every dwelling selected for the Māori booster sample (see Chapter 6 for more details).

2009 NZCASS

In 2009, up to 16 additional dwellings were sampled for the Māori booster sample in each of the 1,000 sample meshblocks in addition to the main sample dwellings. They were selected following the random route method described below. The number of dwellings to approach for the booster sample was held constant at 16, rather than being recalculated for each meshblock. An exception to this was where meshblocks had a low number of dwellings overall, and it was therefore not possible to approach 16 Māori booster sample dwellings.

In these dwellings, residents of Māori ethnicity were eligible for selection. The sampling interval applied to these dwellings was 1, once all the main sample (core) dwellings had been identified and set aside. Starting from the dwelling adjacent to the first selected main sample dwelling, the dwellings 'in-between' the main sample dwellings were consecutively selected, up to a maximum of 16 dwellings.

As a maximum of 26 dwellings could be approached (10 in the main sample and 16 in the Māori booster sample), in the small percentage of meshblocks that contained fewer than 26 dwellings, it was possible that fewer than 16 dwellings would be approached for the Māori booster sample. In practice, often 7 main sample dwellings were approached, as a main sample cluster size of 6.5 was targeted. Meshblocks which contained fewer than 9 dwellings, according to 2006 Census counts, were not included in the sample. In meshblocks with a census count of 9, only main sample dwellings were selected. Thirty-six of the 1,000 meshblocks selected had a census count of 9, and therefore no dwellings were approached for the Māori booster sample.

2006 NZCASS

In 2006, 320 NAUs were selected for the Māori booster sample, and these were separate from the 800 main sample NAUs. Māori booster NAUs were selected with probability proportional to the estimated number of Māori dwellings. An average interview cluster size of 5 was targeted in these 320 NAUs, with the goal of achieving 1,600 interviews. NAUs with a low Māori density were deleted from the sampling frame. As in the 2001 survey, NAUs where less than 5% of dwellings contained Māori were removed from the sampling frame for the booster sample. This accounted for 3% of NAUs, but only 0.2% of Māori households.

Tertiary sampling unit – Respondent selection

To select the respondent within each sampled dwelling, the interviewer asked the person who answered the door for a list of the first names and birth month of every eligible respondent in the dwelling. CBG's Sample Manager software selected the person who had the next birthday to be the respondent. There was no substitution in the case of non-response.

For situations where the next birthday was not known (eg in a household where a flatmate did not know birth months of other flatmates), an alternative procedure based on the alphabetical order of first names was employed. This alternative procedure selected the adult (aged 15 or over) whose first name began with the letter earliest in the alphabet.

Because many types of victimisation are household-based, only 1 respondent per dwelling was selected. This provided efficient measurement of household victimisation, and avoided potential contamination effects that may have arisen if more than 1 person in a household was interviewed. As discussed in Chapter 10, weights for person-based estimates incorporated the number of residents aged 15 or older in each household to remove any household size biasing effect, which is a routine statistical procedure for household-based surveys.

Probabilities of selection

Meshblock

Meshblocks were selected with probability $1{,}000\ dm/D$, where $d_m$was the number of occupied private dwellings in meshblock m according to the 2013 Census, and D was the total number of occupied private dwellings in the sampling frame according to the 2013 Census.

Dwelling

The probability $p_{i,1}$ that each occupied private dwelling was selected for the main sample was

$$\begin{aligned} p_{i,1} &= P(\text{household } i \text{ selected for main sample}) \\ &= P(\text{meshblock } m \text{ selected}) P\left[\begin{array}{l}\text{household } i \text{ selected for main sample} \\ \text{within meshblock } m\end{array}\right] \\ &= \frac{1{,}000 d_m}{D} \frac{1}{x_m} \end{aligned}$$

whether they were on the combined PAF/electoral roll list or were enumerated later.

The probability $p_{i,2}$ that an address was selected for the Māori booster sample would be zero if it was not listed on the electoral roll as containing any electors of Māori descent. Otherwise it was

$$\begin{aligned} p_{i,2} &= P(\text{household } i \text{ selected for booster sample}) \\ &= P(\text{meshblock } m \text{ selected}) P\left[\begin{array}{l}\text{household } i \text{ selected for booster} \\ \text{sample within meshblock } m\end{array}\right] \\ &= \frac{1{,}000 d_m}{D} \frac{0.6 b_{1,m} + 0.4 b_{2,m}}{A_m} \end{aligned}$$

where A_m is the number of addresses listed as containing electors of Māori descent within meshblock m, not counting addresses that had been selected for the main sample; $b_{1,m}$ is the number of addresses that would be selected for the booster sample in meshblock m if it was scheduled for interviewing in February to April; and $b_{2,m}$ is the number of addresses that would be selected for the booster sample in meshblock m if it was scheduled for interviewing in May or June.

The combined selection probability for household i was:

$$\begin{aligned} P_i &= P(\text{household } i \text{ included}) \\ &= P(\text{household } i \text{ selected and eligible}) \\ &= \sum_{s-1}^{2} p_{i,s}\, r_{i,s} \end{aligned}$$

where $p_{i,s}$ is the probability that household i was selected for sample s and $r_{i,s}$ is the probability that household i was eligible for sample s. (The main sample and Māori booster sample are indexed by $s = 1$ and 2 respectively.)

The second element of the combined household selection probability formula is the probability that a household was eligible for each sample. All households in the dataset were eligible for the main sample – that is, the probability of eligibility for the main sample $r_{i,1}$ is always 1. Household eligibility for the Māori booster sample is determined by whether there were any Māori aged 15 or over living in the household – that is, $r_{i,2}$ is 1 if the household contains any Māori aged 15 or over, and is zero otherwise. An alternative estimate of $r_{i,2}$ was used in the calculation of comparison weights, as described in Chapter **Error! Reference source not found.**: Weighting.

Respondent

The probability that respondent i was selected was:

$$O_i = P(\text{person } i \text{ was included})$$

$$= \sum_{s-1}^{2} \frac{p_{i,s} r_{i,s}}{e_{i,s}}$$

where $e_{i,s}$ is the number of people aged 15 or over living with respondent i who were eligible for sample s. The summand for $s = 2$ is taken as zero for non-Māori respondents.

Process for incorporating address files

The process for incorporating the PAF and electoral roll addresses is shown in Table 2.12.

Table 0.1: Process for incorporating address files

Step	Purpose	Process description
1	Create list from which to select addresses	Add addresses from the electoral roll (where an elector of Māori descent resides) to the PAF, if these addresses were not already included in the PAF.[24]
2	Prepare the sample data	Remove incomplete and ineligible addresses from the combined file.
3	Main sample selection	Select addresses for the main sample systematically from the combined list by applying the specified main sample skip interval for each meshblock. (Within each meshblock, addresses were ordered by street address then by street number. A random house was selected in the meshblock, then every k^{th} house from there was selected, where k was the specified skip interval for the main sample in that meshblock.)
4	Prepare the booster sample data	Remove any addresses already selected for the main sample.
5	Māori booster selection	Select a specified number of addresses for the Māori booster sample systematically from the remainder of the electoral roll by calculating and applying a booster sample skip interval, beginning at a random house. (Addresses were ordered by street address then by street number. A random house was selected in the meshblock, then every x^{th} house was selected, where x was the booster sample skip interval for that meshblock.)
6	Enumerated addresses added	Systematically select freshly enumerated addresses (ie any enumerated addresses that did not appear in the combined PAF/electoral roll list) using the main sample skip interval. (The PAF contained no addresses for 1 of the selected meshblocks, despite the census showing occupied private dwellings in that meshblock.)[25]

[24] Addresses where an interview was conducted for the NZCASS pilot study in 2013 were excluded from the combined list.
[25] Only 1 meshblock in the 2014 main study sample contained no PAF addresses.

3. Questionnaire and incident selection

Introduction

The structure and core content of the NZCASS questionnaire has remained consistent since 2006 to ensure key results are comparable.

The main changes made to the questionnaire since 2006 are listed in Table 3.1.[26]

Table 0.1: Questionnaire changes since 2006

Year	Main change	Rationale
2009	Removal of 'electronic crime (e-crime)' and 'cost of crime' question modules.	
2014	Use of Statistics NZ standards for selected demographic questions. Inclusion of 'Security' module.	Bring demographic questions more in line with Statistics NZ standards where possible. Better accommodate stakeholder needs.

Mode of interviewing

Interviews as part of the NZCASS are conducted using:

- computer-assisted personal interviewing (CAPI), where interviewers enter respondents' answers into a laptop.
- computer-assisted self-interviewing (CASI), where respondents are handed the laptop and can enter their own responses.

There are 3 key advantages to this mode of interviewing in relation to the NZCASS:

- computer-assisted interviewing software ensures that survey logic is adhered to
- the selection of the CAPI victim forms can be automated
- respondents can answer sensitive questions confidentially using CASI and reduce bias.

While this mode of interviewing has remained consistent since 2006, the survey software used has changed with each subsequent fieldwork provider (Table 3.2).

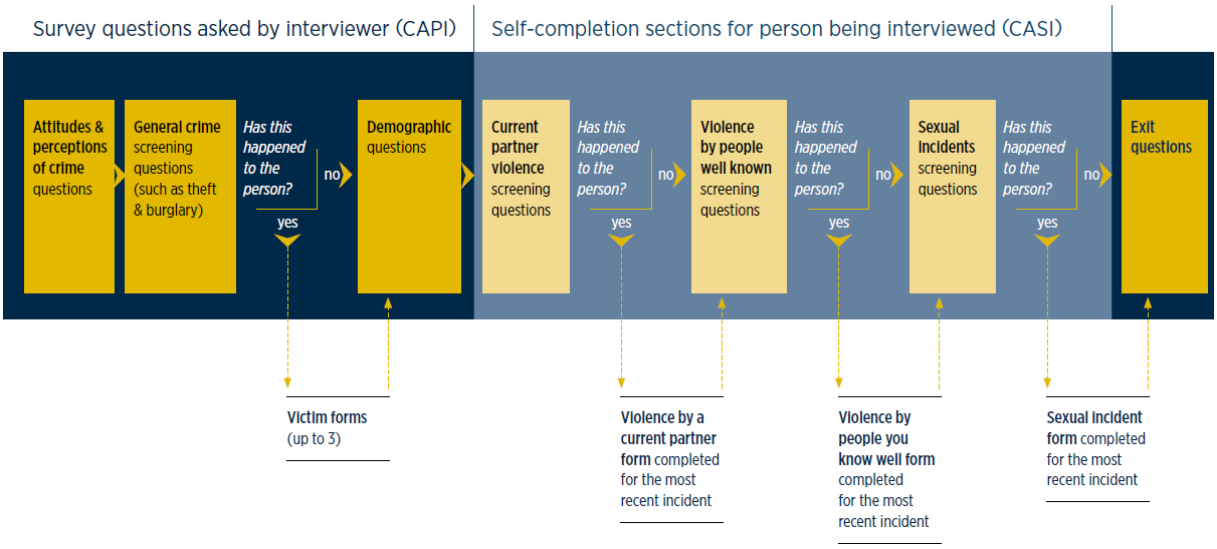
Table 0.1: Fieldwork provider and survey software, by year

Year	Fieldwork provider	Survey software
2006	The Nielsen Company (AC Nielsen)	Confirmit
2009	National Research Bureau (NRB)	Blaise
2014	CBG Public Sector Surveying (CBG Health Research Ltd)	The Survey System (TSS)

[26] It should be noted that this table only presents the main changes to the questionnaire in each year. Other changes have been made to ensure relevance to stakeholders or based on recommendations from previous iterations; however, not all are listed here.

The questionnaire

Error! Reference source not found. **Overview of the structure and content of the NZCASS**



uestionnaire.

Table 3.3: Outline of topics covered in NZCASS questionnaire by section

	Section and questions	
	Interviewer-administered (CAPI)	
Attitude and perceptions	• attitudes to local crime and incivilities • fear and worry about crime • security and neighbourhood support • confidence in the criminal justice system	
Victim screener questions (CAPI screener questions)	• household and personal offences screener questions (excludes violence by partners, people well known and sexual violence)	
General victim forms	• date of offence • same/series of offences • location of offence • mode of entry • contact with the offender • items stolen • damage • insurance	• attempted theft • use of force • threats and weapons • medical attention • emotional reactions • reporting to Police • victim needs** • perceptions of seriousness of incident
Demographics	• age group • ethnicity • employment status**** • marital status**** • hardship****	• income*** • household composition**** • household size • tenure

*New question(s) in 2009; **Modified in 2009; ***New question(s) in 2014; ****Modified in 2014

	Self-completion (CASI)
Partner violence screener questions	• assault • threats of assault • damage to personal property • threats of damage to personal property
CASI victim form 1 (SC1): Partner violence	• reporting to Police • Police response* • victim needs** • emotional reaction • psychological abuse
Violence by 'people well known' screener questions	• assault • threats of assault • damage to personal property • threats of damage to personal property
CASI victim form 2 (SC2): People well known	• reporting to Police • Police response* • victim needs** • emotional reaction
Sexual violence screener questions	• sexual violation (rape) • attempted sexual violation (attempted rape) • distressing sexual touching • other sexual violence or threats
CASI victim form 3 (SC3): Sexual violence	• reporting to Police • Police response* • victim needs** • emotional reaction
Exit questions	

*New question(s) in 2009; **Modified in 2009; ***New question(s) in 2014; ****Modified in 2014

Selection of incidents

At 4 points during the interview (each set of screener questions) respondents are asked how many incidents of each type of crime they experienced (since 1 January 2013). As illustrated above in Table 3.3, respondents are then asked for more detail about some of these incidents via a victim form.

Due to the time it takes for a respondent to complete a victim form, it is not feasible for a heavily victimised respondent to fill in a victim form for each incident they experienced. As such, an automatic incident selection process was developed in 2006 and has continued unchanged through the 2009 and 2014 iterations of the NZCASS.

Selection of incidents in the CAPI section

If a respondent recorded 3 or more incidents in the CAPI screener questions, victim forms were completed for a maximum of 3 incidents which were randomly selected by the computer-assisted interviewing software. If a respondent recorded 3 incidents or less, victim forms were completed for each incident.

The sample design for selecting incidents aimed to:

- ensure the accuracy of incidence and prevalence rates for key offence types
- provide sufficient victim form information on the characteristics of major offence types
- maintain consistency with the approach used in the 2006 and 2009 surveys.

Method:

1. If a respondent reported having experienced 3 or fewer incidents across all CAPI screener questions, a victim form was completed for each of these incidents. If the respondent reported experiencing more than 3 incidents at the CAPI screener questions, all these incidents were placed into the 'pool of incidents' experienced by that respondent, and from this pool, the 3 incidents for which victim forms would be completed were randomly selected as described below.
2. Each incident was assigned a selection weight according to the screener question at which it was reported. Screener questions fell into 3 priority categories (low, medium and high), with corresponding selection weights (1, 2 and 3) as shown in Table 3.4.
3. Incidents were selected independently, without replacement, with selection probabilities proportional to the weight given to the incident's screener question. In other words, the selection weight for incident i was denoted by w_i. Then the probability of selection for incident j for a particular victim form was $w_i/sum\ (w_i)$, where the sum was taken over all incidents available for selection at that stage. Incidents were selected without replacement, so those that had already been selected for an earlier victim form would not be included in the sum.

The probability of selection for a particular incident depended on both:

a. the extent of competition from other incidents

b. the screener question that the incident was recorded at.

Once the incident selection design and method was implemented in CAPI, quality assurance testing took place to ensure that it worked as intended.

Each of the 15 incident types (based on screener questions) was assigned a number and weight as shown in Table 3.4.

Table 3.4: Incident selection weight level

Incident number	Incident description	Question number	Weight level
1	Taking/conversion motor vehicle	28	3
2	Theft from motor vehicles	29	1
3	Damage to motor vehicle	30	1
4	Attempt to break into home/garage	31	2
5	Unlawful entry into your home/garage	32	1
6	Theft from outside the home (over $10)	34	1
7	Theft from inside the home (right to be there)	35	2
8	Wilful damage to household property	35.416	2
9	Assault	36	3
10	Threatening to assault	37	3
11	Wilful damage to personal property	38	3
12	Threatening to damage property	39	2
13	Theft/attempted theft from person	40	3
14	Theft of personal property	41	2
15	Other offence type	43	1

Selection of incidents in the CASI section

As described above under 'The questionnaire', the CASI section of the questionnaire covered incidents that were of a more sensitive nature than those covered in the CAPI section. The CASI section contained 3 victim forms: violence by a partner; violence by people well known to the respondent; and sexual violence.

Method:

1. Four screener questions preceded the victim form in each part of the CASI section.
2. If a respondent answered 'yes' to at least 1 of the screener questions in that part, they then completed that part's victim form.
3. Where there was more than 1 incident reported within that CASI section part, the respondent was asked to think about the most recent incident and complete a victim form for that incident.

Some of the heaviest victimisation was recorded in the CASI section. Because only 1 victim form was allocated to each part, the probability of selecting each incident experienced by heavily victimized respondents was very low. This resulted in highly variable incident weights, and may mean that the incidents with missing data are not similar to other incidents.

Incident selection testing

The selection of incidents is a critical part of the NZCASS and is needed to ensure that national victimisation incidence and prevalence rates can be accurately estimated.

To ensure that the incident selection process was working correctly within the survey software, a testing process was designed and assessed by Statistical Insights and implemented by CBG prior to the start of the pilot study fieldwork. Over 250,000 simulated interviews were run across a range of scenarios to check that the correct selection probabilities were being used.

The testing procedure comprised the following steps:

1. The 12 sets of values were programmed into the survey software so that particular answers were given to certain screener questions for each of 12 different scenarios.
2. Up to 3 incidents were selected for the victim forms.
3. The survey software was programmed to create batches of at least 5,000 synthetic interviews for each set of values (answers). Each of the synthetic interviews showed which incidents were selected.
4. Varying numbers of synthetic interviews were created depending on the likelihood of respondents experiencing the incidents referred to by each set of data values. This was done to ensure that incident types which occur far less often than others have a chance of being selected, and thus the programming could be checked.
5. Discrepancies between the expected incident selection probabilities and the simulated interviews were then assessed by Statistical Insights to ensure the incident selection process was working correctly.

4. Questionnaire review, development and testing

Questionnaire review and development

The NZCASS questionnaire was reviewed between June and August 2013. Although some changes were made to the questionnaire used in 2014, it was based on and very similar to those used in both the 2006 and 2009 iterations of the NZCASS.

During the questionnaire review and development process the project team was mindful of the following key points:

1. The need to ensure that the core statistics collected as part of the NZCASS (eg incidence and prevalence) remained comparable with previous iterations of the survey. This meant that even a seemingly small change to a screener question may have an unintended effect on our ability to look at trends.
2. The NZCASS is an extremely long questionnaire with a number of complex routing and conditional programming requirements. As such, changing 1 part of the questionnaire can easily lead to unintended consequences elsewhere.
3. Different parts of the questionnaire are used by different stakeholders in a variety of ways. This means that each question proposed for review needed to be checked against (known) user requirements to assess the impact of changing it.

The review process comprised the following steps:

1. Each question that inputted into 1 of the following processes was highlighted to ensure that they were not inadvertently tampered with during the review process:
    a. incident selection
    b. offence coding
    c. data checking/validation processes
    d. weighting or imputation processes
    e. key reporting requirements.
2. At the conclusion of the 2009 NZCASS, a number of suggestions were made for the next project team in relation to questions that should be reviewed and potentially changed, removed or updated. This list of recommendations was the starting point for the 2014 review process with each suggestion being reviewed and assessed for its relevance and feasibility in 2014.
3. Where possible, stakeholder information needs were incorporated into new or existing questions.
4. Each question that was reviewed or proposed was entered into the 'Questionnaire Review Register' and assessed for the impact that the proposed change would have on:
    a. the research objectives
    b. trend data/continuity
    c. indicators/measures used from the NZCASS
    d. stakeholder needs

e. timing and cost

f. routing implications

g. offence coding

h. weighting and imputation processes

i. previous analysis and reporting products.

5. Where changes were proposed that had an impact on 1 or more of the elements above, these were discussed with the wider project team and approved by the NZCASS Steering Group prior to the change being implemented.

Cognitive testing

Cognitive testing was undertaken by CBG on 21 new or existing questions in order to:

- check participants' comprehension of wording used
- check participants' understanding of the concepts associated with each question
- understand how participants recalled information relating to each question
- understand how participants made response decisions for each question.

Table 4.1: Key elements of the cognitive testing process

	Details
Testing period	24–25 August 2013
Sample	Sixty interviews were undertaken in Auckland and Taupo; 48 interviews were conducted in Auckland with the remainder in Taupo. Auckland and Taupo were chosen as test locations due to the ethnic diversity of participants within the recruitment networks.
Recruitment	Participants were specifically recruited via interviewers' existing professional and personal networks.
Questions tested	Twenty-one new and existing questions were tested including the informed consent introduction for data linking.
Who conducted the interviews	Four researchers experienced in undertaking cognitive tests conducted the interviews. Each researcher undertook 15 interviews.

Table 4.2 details the number of interviews completed by demographic group. The sample consisted of 24 males and 36 females.

Table 4.2: Cognitive testing number of interviews completed by ethnicity and age

Age group	Māori	Pacific	Asian	NZ European	Other	Total
15–21	4	5	1	9	0	19
22–45	5	4	4	7	2	22
46+	7	4	1	7	0	19
Total number of interviews	**16**	**13**	**6**	**23**	**2**	**60**

Based on findings from the cognitive testing process, some minor adjustments were made to some of the new questions being proposed for inclusion (eg new security questions) and the data linking introduction, which aims to give an overview of the data linking process and seek informed consent for this process.

No further rounds of formal cognitive testing were possible due to time constraints.

CAPI/CASI programming and testing

Due to project time constraints, questionnaire programming and testing was conducted as a parallel process alongside the development of the pilot study questionnaire. While not ideal, this was necessary in order to meet the required timeframes.

As noted above, new CAPI/CASI software was used in the 2014 NZCASS. As programming progressed, a number of limitations were discovered in The Survey System (TSS) software, particularly in relation to some of the more complex conditional questions and routing structures. To compensate for these limitations, additional programming had to be coded outside the software for a number of questions in order to achieve the desired functionality and behaviour.

Throughout both the programming process and prior to the pilot study, the questionnaire was tested by a professional software tester at CBG. Checks included (but were not limited to):

- question and response text matched the supplied questionnaire document
- multi/single response questions allowed multiple and single responses as applicable
- response ranges were within the boundaries defined by the survey
- text could be entered for questions allowing free-text ‘Other’ responses
- all previously entered response options were removed when the ‘reset answers’ button was selected
- unique responses could not be selected along with other responses in multiple choice questions (eg you shouldn’t be able to select ‘Don’t know’ along with any other response options)
- skip instructions worked correctly for questions with ‘go to’ instructions
- where a question had no skip instructions, all response options were checked to ensure they went to the next question
- logic test cases were executed
- the ability to go back through the questionnaire to make corrections to previous entries was also tested.

As part of the testing process, ministry personnel (predominately in the Research & Evaluation team) also tested an online CAPI/CASI version of the questionnaire extensively and worked with CBG to find and resolve issues.

Programmed checks

In addition to the manual checks noted above, a range of logic, consistency and range checks were programmed into TSS code to ensure the data was correct and robust. Checks can be categorised as follows:

- **hard error checks:** required interviewers to change data that they had entered
- **soft error checks:** gave the interviewer the opportunity to check and possibly change the data they had entered.

The 3 main types of checks conducted were:

1. logic checks
2. consistency checks
3. range checks.

Logic checks
This type of check is commonly applied in multiple choice questions where a list of response options is given along with a 'non-response' option (eg 'Don't know' or 'Refused') and where that non-response option is considered 'a unique code' (ie cannot be selected together with any of the other responses).

For example, 1 of the security questions asked respondents for the reasons why they had not participated in a neighbourhood support group in the last 12 months. There were 2 answers within the response framework which could not be selected in conjunction with any other answer: 'Not aware of such a group in my area' and 'Don't know'.

Consistency checks
This type of check is used in questions where it is important to ensure consistent information is entered – for example, in the demographic questions where it is important to ensure consistency between the number of people in the household, the number of children in the household and the household composition characteristics.

Range checks
For some questions, the data entered has to be within a certain range. Range checks prompted interviewers (or respondents in the CASI section) to change their answer where an answer outside of the acceptable range had been entered.

For example, the numerical range for all of the CAPI victimisation screener questions was 0 to 97.

Change control process

Throughout the CAPI/CASI testing process, a working register of all issues, discussions and resolutions was maintained by CBG and the ministry. This register is now held with the ministry as part of the project record.

The pilot study commenced only once all issues that impacted the collection of robust data were resolved.

Pilot study

The purpose of the pilot study was to mimic the main study as closely as possible, to ensure that the questionnaire and associated survey processes were robust and functioning correctly. In 2014 both fieldwork and the offence coding activities were undertaken by the same provider (CBG). As such, the pilot study tested both of these work streams.

In particular, the pilot study aimed to ensure that:

- new or altered questions were tested to ensure they were understandable, working as intended, interviewer prompts were suitable and response frameworks were appropriate
- questionnaire routing was working correctly
- the questionnaire loaded into the CAPI and CASI software worked correctly, as well as the electronic sample management
- interviewer and coder training and resources were fit for purpose
- fieldwork and coding reporting and monitoring systems/processes were functioning as expected
- fieldwork and coding statistics were as expected and/or within scope of project costs, time and quality parameters
- incident selection worked as expected (secondary check)
- fieldwork communications were appropriate by gathering high-level feedback from respondents
- any risks to the main study were identified and suggested responses provided.

Table 4.3: Methodology and key information

	Details
Overview	Most methods and processes used as part of the pilot study were in line with those planned as part of the main study. The main difference between the pilot and main studies was the meshblock sampling process. Meshblocks used as part of the pilot study were chosen (rather than randomly selected) to: • provide a mix of urban and rural areas • ensure that high crime areas were over-represented. This was done to maximise the probability of encountering respondents who had experienced crime and test both the victim screening process and the questions asked in the victim forms.
Target population	Total usually resident, non-institutionalised, civilian population of New Zealand aged 15 years and over.
Sample design	Twenty-five meshblocks were sampled, 10 of which (40%[27]) were classified as 'high-crime areas'.[28] All high-crime areas were also highly deprived (New Zealand Index of Deprivation (NZDep) quintile 5). As with the 2008 pilot study, 15 medium/low crime areas were taken from NZDep quintiles 1–4.

[27] In 2009, 41% of meshblocks were selected from high-crime areas.
[28] Defined as Police station areas with high offence rates per capita.

	Details	
Sampled areas	Auckland, Hamilton, Wellington, Gisborne.	
Dwellings included	Permanent, private dwellings.	
Sample size	Main sample:	102
	Māori booster sample:	69
	Total sample:	**171**
Response rates	Main sample:	56.9%
	Māori booster sample:	73.6%
	Total sample:	**62.6%**
Interviewing period	18 November – 1 December 2013	
Average interview length	Total questionnaire	42 minutes and 44 seconds
Questionnaire recall period	1 January 2013 to date of interview	

Interviewers and training

Six interviewers were involved in the pilot study and were trained over a 2-week period consisting of a period of mentored self-directed learning and culminating in a 2-day workshop in Auckland.

Prior to attendance at the workshop, interviewers were required to complete a preliminary training course which provided them with foundational knowledge on the survey, experience using the NZCASS electronic Sample Manager and familiarity with the questionnaire. As this knowledge was assimilated in advance, the workshop programme could focus on the more practical aspects of the fieldwork and also meant that time was available for guest speakers to present to the group.

For more information on interviewers and interviewer training please refer to Chapter 5: Fieldwork processes.

Post-pilot changes

Only 2 material changes were made to the questionnaire based on recommendations from the pilot study:

1. Two response options were added to 1 of the new 'security questions'.
2. The 'Don't know' response option was removed from the showcard relating to the personal and household income questions. This was done in an attempt to avoid high rates of non- response to these questions.

5. Fieldwork processes

Introduction

Key fieldwork processes in 2014 remained largely the same as in 2006 and 2009 to help retain consistency and ensure comparability between years. Where processes changed, this was predominantly done to help improve response rates and increase efficiencies in fieldwork management.

Fieldwork period

The fieldwork period for the 2014 NZCASS aimed to mirror that of 2009 as closely as possible. As can be seen from Table 5.1, the start and end dates for 2014 fieldwork are comparable.

Table 5.1: Fieldwork period, by year

Year	Period	Notes
2014	Monday 10 February 2014 to Sunday 6 July 2014	All required contacts were made by 30 June 2014 with the final week up until 6 July being used only to complete outstanding appointments and other 'mop-up' activities.
2009	Saturday 14 February 2009 to Sunday 5 July 2009	• Fieldwork period includes enumeration, household selection, setting up interview appointment times and interviewing. • The fieldwork period was 142 days in total. • Interviews were conducted over 139 days from Tuesday 17 February to Sunday 5 July.
2006	Thursday 9 February 2006 to Sunday 25 June 2006	First year of the NZCASS.

Issuing meshblocks

One thousand meshblocks were divided and allocated to each of the 5 fieldwork months. Meshblocks were progressively issued to interviewers as fieldwork advanced.

Table 5.2: Month of issuing meshblocks - 2014

	Feb	Mar	Apr	May	June	Total
Meshblocks issued	204	201	200	197	198	1,000
Estimated eligibles	1,713	1,687	1,662	1,761	1,744	8,574

Interviewers and training

Table 5.3: Overview of interviewers and training

	Notes
Interviewers	Twenty-six interviewers were selected from a pool of experienced CBG interviewers who had a proven track record working on other large government surveys. No new interviewers were used as part of the 2014 NZCASS field team.
General interviewer skills and training	All interviewers had completed the following CBG training modules: • public sector surveying • maximising response rates • cultural awareness • enumeration • safety management. Since all interviewers working on the 2014 NZCASS were established and experienced CBG staff, all these modules had been completed prior to the start of NZCASS training.
Pre-reading and study	Prior to the NZCASS training days, interviewers received a copy of the NZCASS questionnaire, the survey manual and the NZCASS workbook. Interviewers were expected to spend time prior to the NZCASS training days studying this material and becoming familiar with NZCASS interviewing processes.
Practice	As part of the NZCASS preliminary training, interviewers were required to practice administering the NZCASS survey on friends/family so that they could practice the questionnaire and help to become familiar with its application and layout.
NZCASS training days	As part of the NZCASS main study, all 26 CBG interviewers attended 2 days of training in Auckland on the 4th and 5th of February 2014. These days consisted of (but were not limited to): • introduction and background to the NZCASS • presentations by the ministry (NZCASS Project Manager) and New Zealand Police on the use and importance of the NZCASS • a presentation by Victim Support discussing victims experiences, reactions and needs • recruitment for the NZCASS, including strategies to maximise response rates and overcome reluctance • the questionnaire, with a focus on the most important questions, the more complicated parts of the questionnaire, and things to note. There was also a specific focus on the purpose of different questions so that interviewers had more understanding of why certain questions were asked and where the questionnaire differed from what they might be used to in their application of other surveys. • discussion of situations that might be encountered during the application of the NZCASS and different ways to handle these (eg if a respondent is fearful of answering due to the presence of a family member, or if a respondent becomes upset due to the nature of the questions)

	Notes
	• fieldwork processes for the NZCASS with special attention paid to processes that interviewers might not be used to through their application of other surveys (eg NZCASS contact and outcome codes) • audit and quality assurance processes that would be employed during the fieldwork • overview of the Specialist Offence Coding process. The NZCASS Project Manager and a senior advisor from the ministry attended both training days.
Assessment	In preparation for fieldwork, all interviewers were assessed by the CBG managers to confirm that they were ready to begin interviewing as part of the NZCASS. Interviewers were not permitted to begin interviewing as part of the NZCASS until they had completed all the required training, undertaken the required practice interviews and passed the assessments.

Fieldwork resources

Interviewer resources

NZCASS interviewers were provided with a number of resources to assist them during the fieldwork period. Table 5.4 provides a summary of these resources.

Table 5.4: Interviewer resources

Resource	Description
Survey manual	The manual is designed as a step-by-step guide to survey administration and includes (but is not limited to): • information about the survey • workflow (meshblock) planning • using the Sample Manager software • enumeration • participant selection • informed consent • contact outcome coding • the questionnaire • closing the survey • administration.
NZCASS workbook	The workbook provides contextual information, information about the survey's content and design, and a questionnaire study guide. This resource also provides detailed training about signs of respondent distress when dealing with sensitive content, how to build rapport and put the respondent at ease and encouraging responses to the self completion section of the questionnaire.
Showcards	Showcards were provided in booklet form to all interviewers.
Consent forms	Consent forms in 2014 were in electronic format on the interviewing laptops. Respondents signed electronically using their finger or a stylus to record consent. Paper copies of the consent forms were left with respondents for future reference
Tablet/laptop	New technology replaced paperwork used in previous iterations of the NZCASS with sample management and respondent selection taking place within CBG's Sample Manager software. Electronic copies of meshblock maps and participant information sheets were also incorporated into the programme.

Respondent resources

As shown in Table 5.5, a number of fieldwork resources were produced as part of the survey to assist interviewers when engaging households/respondents and to help answer respondent queries.

Table 5.5: Respondent resources

Resource	Description
Letter to household	A letter was sent from the ministry introducing the survey, introducing CBG as the ministry's fieldwork provider, and encouraging participation when the interviewer visits. The letter was sent out to households in batches 7–10 days before the interviewer was due to call. This was done in order to improve householders' recall of the letter. Interviewers were also given spare copies of the letter to help engage respondents at the door if they didn't remember receiving it in the mail.
Information pamphlet	A professionally designed pamphlet containing key information about the NZCASS was also mailed with the letter, including (but not limited to): • what the NZCASS is and why we do it • what the information collected is used for • what type of questions are asked • who conducts the NZCASS and when it will be undertaken • who will be asked to participate • 0800 numbers for both CBG and the ministry, should participants want to confirm the validity of the research or talk to someone further about it.
Thank-you card	At the end of the interview, a thank-you card was offered to participants. The thank-you card contained contact details for a range of support organisations that provide assistance to victims of crime.

Table 5.6 provides a summary of the fieldwork products produced and used during the main study.

Table 5.6: Number of fieldwork products used

Product	Produced	Dispatched	Balance at end of fieldwork	Total used
Letter to household (English)	14,234	11,506	52	14,182
Letter to household (Māori)	500	130	451	49
Information pamphlet (English)	18,500	18,317	1,089	17,411
Information pamphlet (Māori)	500	260	434	66
Thank-you card (English)	7,500	7,500	1,064	6,436
Thank-you card (Māori)	500	260	391	109

Translations

The fieldwork resources above were translated into a number of languages in 2014. Interviewers had translations of both the 'Letter to Household' and 'Information Pamphlet' available on their laptops for use when needed. These products could be printed out by interviewers and left with survey participants as and when needed but only English and Māori products were printed in brochure format.

In 2014, the fieldwork provider monitored translation/interpreter requests from respondents. A summary of the languages and frequency of requests is provided in Table 5.7.

Table 5.7: Number of translation requests

Language	NZCASS translated information available	Translation/interpreter requests
Māori	Yes	0
Niuean	Yes	0
Fijian	Yes	0
Samoan	Yes	5
Tongan	Yes	1
Cook Island Māori	Yes	1
Chinese (simplified)	Yes	0
Chinese (traditional)	Yes	0
Korean	Yes	1
Hindi	Yes	1
Other – total	–	21
Other – Mandarin	No	9
Other – Serbian	No	1
Other – Vietnamese	No	5
Other – Japanese	No	1
Other – Burmese	No	1
Other – Cantonese	No	2
Other – Croatian	No	1
Other – Bahasa (Indonesian)	No	1

Fieldwork procedures

Table 5.8: Fieldwork procedures

	Details
Visiting days and times	Interviewers approached households 7 days a week between the hours of 9:00am to 8:00pm. Occasionally, respondents requested an appointment time outside of these hours with the interviewer accommodating wherever possible. In order to increase the likelihood of finding a resident at home, interviewers visited households on a mixture of weekdays and weekends and at different times of the day. There were no differences in visiting days or times between urban and rural areas.
Visits to meshblocks	Each meshblock was visited by an interviewer a minimum of 5 times[29] unless the interviewer had achieved or recorded a final contact outcome for all selected households in a meshblock. Interviewers were allowed to visit each meshblock on 7 different days, although multiple visits on the same day were allowed. Typically, trips to each meshblock were spread over an average of 4 weeks.
Call-backs	Up to a maximum of 9 call-backs (10 calls in total) were made to selected dwellings.
Electronic sample management	All fieldwork activity was recorded in CBG's Sample Manager software installed on the laptop computer of each interviewer. The software contained records for every selected house in the sample and provided the ability to perform respondent selection at the door according to survey protocols. The Sample Manager also provided the interviewer with access to meshblock maps, translated fieldwork products and links to launch the survey.

[29] The procedure used in 2006 to visit areas and approach dwellings was different – refer to Reilly and Sullivan (2008: 12).

Fieldwork management

A number of processes were put in place to ensure that interviewers were supported through the fieldwork process and interviewing was completed on time and to the required standard.

Interviewers were monitored during fieldwork by the CBG field management team. Survey completion rates and data quality were examined regularly at the individual interviewer level to ensure that all interviewing was completed within the required timeframe and to a high quality.

Interviewers attended weekly teleconference meetings where the survey management team communicated key messages and shared learnings. The meetings were also used to discuss overall progress and celebrate successes. Each interviewer was also able to monitor their own progress and performance throughout the fieldwork via their own personal web portal. Where it was identified that an interviewer required additional training or support, this was provided.

Fieldwork progress, monitoring and reporting

As part of monitoring practices and reporting to the ministry, an online dashboard was set up by CBG so that fieldwork statistics could be viewed in real time by project staff. In addition to this, a formal monthly fieldwork report was provided to the ministry as a summary of progress and as part of the project record.

Table 5.9 provides an overview of cumulative number of interviews (targeted and completed) throughout fieldwork.

Table 5.9: Number of interviews targeted and completed, by month

	Feb	Mar	Apr	May	June	Total
Targeted number of interviews (cumulative)	1,292	2,584	3,876	5,168	6,460	6,460
Total number of interviews completed (cumulative)	1,025	2,245	3,959	5,183	6,832	6,943[30]
Percentage complete	15.9%	34.8%	61.3%	80.2%	105.8%	107.5%

[30] Final total includes last interviews 'mopped up' in the first week of July 2014.

Fieldwork lessons

As noted above, interviewers attended weekly teleconference meetings where the survey management team communicated key messages and shared learnings. Table 5.10 outlines the main lessons learnt during the 2014 main study fieldwork.

Table 5.10: Fieldwork lessons

Lesson	Details
Invitation letter	Sending an invitation letter and information pamphlet to selected houses advising the household of the impending interviewer's visit produced a positive response. Most respondents had seen the invitation letter and were better informed when the interviewer visited.
Data linking	During fieldwork it was identified that some interviewers were not consistently recording respondents' surnames for the purpose of data linking. Additional training was provided to the field team and data collection was monitored to ensure high rates of surname collection were maintained for the remainder of the project.
Q47 free-text responses	It was identified in 2009 by the coding provider at the time that some of the free-text incident descriptions recorded at Q47 were of low quality, making a coding decision difficult. At the beginning of the 2014 fieldwork, the coding supervisor audited a selection of responses recorded at this question by all interviewers. Additional training was then provided to interviewers who were not recording enough detail, or who were not recording the incident particulars in the first person.

Fieldwork quality assurance

A number of quality assurance processes were in place for the 2014 NZCASS fieldwork. These processes ensured that all risks were managed and fieldwork progressed on time and to the required standard. Fieldwork processes were implemented and managed by the ministry's contracted fieldwork provider, CBG.

Overview of fieldwork quality risks

There are a number of risks that can have an impact on the quality of the data collected and potentially the number of CAPI or CASI victim forms completed. Table 5.11 provides a list of some of these risks.

Table 5.11: Overview of fieldwork quality risks

	Risk	Description
1	Interviewers do not visit sampled households as required	The NZCASS sampling process has been carefully designed to ensure that households throughout the country from both low- and high-crime areas are selected. If interviewers do not visit households according to the required sampling process, there is the risk that biases will be introduced which may impact the number of CAPI or CASI victim forms being collected.
2	Incorrect householder sampled	If the required respondent sampling process is not followed, the incorrect person may be selected. For example, if only the people present at the time of visit is entered into the sampling system (rather than all the people living at the address), an incorrect respondent may be selected.
3	Screener questions asked incorrectly	The number of victim forms completed relies on the number of screener questions where a respondent answers affirmatively that they've experienced an incident. The number of victim forms selected can also be affected if the interviewer does not ask the screener questions correctly and insert emphasis on the correct words.
4	Self-completion handover process executed incorrectly	Directly after the demographic section, interviewers are trained to introduce the CASI section of the questionnaire and encourage respondents to participate – even if they haven't experienced a crime. At this point interviewers are asked to enter a response to Q163, which asks whether the respondent is happy to continue or not: 1. happy to continue (computer is passed to the respondent) 2. hesitant (respondent is hesitant to continue but interviewers are trained to reassure them and to help them get started) 3. refused totally (response option only used when the respondent completely refuses and the interviewer can't persuade them to continue with the CASI section). If an interviewer is not skilled at handling respondents' concerns if hesitant and encouraging participation – even if the respondent hasn't experienced a crime – respondents can drop out at this point of the questionnaire and hence the number of CASI victim forms could fall.
5	Respondents exit the interview prematurely	Respondents could end the interview prematurely by not completing the rest of the CASI section and handing back the computer early.

	Risk	Description
6	IT issues occur	There are a number of IT issues that could impact the number of victim forms being submitted. It is up to interviewers to identify if and when these are happening (if they occur during the CASI section, identification will be more tricky) and report them for resolution.
7	Poor response rates and targeted sample not achieved	If a good response rate of the targeted sample size is not achieved, then the number of CAPI and CASI victim forms could be lower.
8	Interviewers falsifying surveys	If interviewers falsify surveys, then the integrity of the data could be compromised.

Quality assurance processes

Table 5.12 lists the main types of processes in place during the 2014 NZCASS fieldwork.

Table 5.12: Quality assurance processes

Process	Description/Purpose
In-field data quality	Monitor key statistics that indicate whether or not surveys are being completed according to the required protocols.
Analysis of survey data	Assess the quality of the data being collected.
Telephone audits	One in every 10 respondents and at least 1 respondent in every meshblock is contacted. A meshblock can't be closed without a successful audit. Audits confirm the following: • the interview took place • the interview took place at the correct address • the number of occupants aged 15 or over living at the address at the point of recruitment • respondent selection procedures were completed correctly including the correct recording of birth month and ethnicity information • the respondent was happy with the way the survey went and with the interviewer • if the respondent had any problems or issues when answering the questions • the respondent completed some questions by themselves using the computer • showcards were used • reason for participation.

Quality assurance – management and statistics

Table 5.13: Fieldwork quality assurance – management and statistics

Interviewers do not visit sampled households as required	
Risk description	**Quality assurance processes**
The NZCASS sampling process has been carefully designed to ensure that households throughout the country from both low- and high-crime areas are selected.	All sampled houses are pre-selected using the NZ Post address database. Selected addresses are pre-loaded into the Sample Manager database used by each interviewer.
If interviewers do not visit households according to the required sampling process, there is the risk that biases will be introduced which may impact the number of CAPI or CASI victim forms being collected.	The Sample Manager will only allow contact, outcome and survey data to be entered into selected address records. This data is uploaded on a daily basis. Data uploaded from the field is used to ensure survey protocols are being followed.

Quality measure	Description	Result	Notes/Comments
Survey completed in the correct address	Respondents are asked during audit telephone calls to confirm that they live at the sampled address where the survey was completed. Ensures that the random sample is protected and the correct houses are surveyed.	95.9%	Occasionally the interviewer will enter data into another sampled address record. Where the respondent reports that the address is not correct, CBG checks to ensure that they indeed live in another sampled house.
Enumerated houses in meshblock	Number of houses that were added into the sample by the interviewer. Expressed as a proportion of the total sample. Results analysed at an individual level to ensure that each interviewer is completing enumeration.	1.6%	

Incorrect householder sampled	
Risk description	**Quality assurance processes**
If the required respondent sampling process is not followed an incorrect person may be selected. For example, if only the people present at the time of visit is entered into the sampling system (rather than all the people living at the address), an incorrect respondent may be selected.	Respondent selection requires the interviewer to list all occupants aged 15+ into the Sample Manager. Month of birth and ethnicity information is also collected. Once all occupants have been added, the Sample Manager automatically selects the person to be approached for the interview based on sampling rules for the survey, thus reducing the possibility of human error resulting in an incorrect occupant being selected. Occupancy information for every household is sent back to CBG where it can be used in further auditing processes/analysis to ensure survey protocols have been followed. This differs from the approach in 2009 where names were listed on paper sampling forms and the selection was made manually by the interviewer.

Quality measure	Description	Result	Notes/Comments
Occupancy match rate	Match rate between the number of occupants recorded in the Sample Manager and the number of occupants reported by the respondent in demographics section of the survey.	92.6%	In 95% of cases the correct number or more occupants were recorded in the Sample Manager. Occasionally, incorrect occupancy information is provided at the door. Or the person completing the survey is not the person who provided the original occupancy information.
Complete month of birth (MoB) information	Proportion of households where MoB information was recorded for all occupants. Respondent selection protocol dictates that the person with the next birthday in the household is selected for the survey. If MoB information is missing for any of the occupants, the person with the first name alphabetically is selected. This check ensures that the selection protocols are being followed.	95.2%	MoB data was collected for all occupants at 95.2% of addresses. All of this information is stored electronically by the Sample Manager and can be used for further quality checks. This audit indicates that in the vast majority of cases, respondent selection is based on the person with the next birthday in the household, rather than the person with the first name alphabetically.
Total adult (15+) occupants recorded	Respondents are asked in the audit telephone call to report the number of people aged 15+ that were living in the household at the time of the interviewer's visit. This measure sums all of the reported occupants from the audit calls and compares the figure to the number of occupants recorded in the Sample Manager for all of the audited houses. This is a high-level check to ensure that occupants in all selected houses are included in the Sample Manager database and have a chance of being selected.	98.8%	There were 2105 occupants reported from all audited houses vs 2079 recorded in the Sample Manager database for these same houses. This check evens out any household-level discrepancies and indicates that almost every eligible occupant in the sampled houses had a chance of being selected.

Quality measure	Description	Result	Notes/Comments
Houses where Sample Manager occupants < audit occupants	The proportion of audited households where the occupants recorded in the Sample Manager was less than the occupants reported by the respondent in the audit call. Seeks to ensure that at a household level, the correct number of occupants is being recorded.	6.6%	The number of occupants recorded in the Sample Manager should match what is reported in the audit call. Occasionally incorrect information is given at the door or the audit question is misunderstood to be asking for total occupants, rather than those aged 15+.
Occupant MoB match	Proportion of occupants with correctly recorded birth month information. Audited occupant is chosen at random from all the occupants listed in the occupant table.	96.7%	This indicates that MoB data is being recorded correctly and not falsified to influence the selection process.
Names and MoB information requested for all occupants aged 15 and over	Proportion of respondent that reported in the audit telephone call that they were asked for this information as part of the selection process. Used to ensure that selection protocols are followed correctly.	96.3%	Rate indicates that this information was consistently requested at the point of recruitment and selection protocols were being followed.
Booster respondents identify as Māori	Proportion of respondents in the booster sample that identify as Māori. Used to ensure that the selected respondent is of eligible ethnicity.	100%	Rate indicates that respondents in booster houses were of eligible ethnicity. Any respondents that respond negatively are subject to a further check which ensures that they selected Māori as 1 of their ethnic groups at Q151 in the survey.
Reselect function use	Number of times the survey respondent is reselected as a proportion of total household selections. Reselection of the respondent is possible in very limited circumstances,[31] when it is later discovered there was some mistake in the data used for respondent selection. The interviewer is required to obtain an authorisation code from survey managers to proceed with reselection. Indi vidual interviewer rates are monitored to ensure that reselection instances conform to survey protocols.	0.85%	Rate indicates that reselection occurs extremely infrequently.

[31] These include when the selected respondent does not usually live in that household, when they are actually under 15 years old, when additional adult occupants are identified after selection took place, when MoB was entered incorrectly, and when a new Māori occupant is identified in a booster sample dwelling.

Screener questions not asked correctly	
Risk description	**Quality assurance processes**
The number of completed victim forms relies on the number of screener questions where a respondent answers 'yes' they've experienced an incident. The number of victim forms selected can also be affected if the interviewer does not ask the screener questions correctly and insert emphasis on the correct words.	Victim form completion rates were monitored closely at the individual interviewer level as low rates may indicate that the interviewer was not administering the screener questions correctly. Those interviewers with lower victim form yields were all subject to a survey deli very assessment focusing on this part of the survey. In all cases, the questions were being administered correctly. Following the concerns raised with low victim form numbers at n = 200 interviews, all interviewers were contacted to confirm that they were not experiencing any issues with the victim form screener questions and incident selection. Specifically, they were asked to confirm that where 1 incident was reported, 1 victim form was chosen for completion; 2 incidents = 2 victim forms; and 3+ incidents = 3 victim forms. All interviewers confirmed that this section of the questionnaire was working in the correct manner.

Quality measure	Description	Actual	Notes/Comments
Household access to vehicle	Proportion of respondents that own or have regular access to car, motorcycle, van or truck. This is victim form screener Q27. If this question is not asked/answered correctly the respondent skips 3 further screener questions relating to vehicle offences with the potential to lose victimisation data. Vehicle-related crime makes up a significant proportion of crime reported in the NZCASS.	91.4%	The rate was 92.5% according to the 2013 Census. Survey results closely match this, indicating that Q28–30 screener questions were not being inadvertently skipped.
Average victim forms completed per survey	Average number of victim forms completed. Designed to identify individual interviewers who may not be completing the screener questions correctly. Indi vi dual rates ranged from 0.2 in Southland to 0.8i n South Auckland.	0.41	The 2009 rate was 0.7. The average number of completed victim forms is dictated by the prevalence of crime during the recall period.
Showcard use	Proportion of respondents that responded affirmatively to the question: 'Did [interviewer] use a book of showcards to help you answer some questions?' in the telephone audit call. Showcards are used throughout the survey to help the respondent answer questions. One card in particular is left visible during the victim form screener questions as a prompt.	95.6%	Rate indicates that showcards were consistently used in field.

Self-completion handover process executed incorrectly	
Risk description	**Quality assurance processes**
Directly after the demographic section, interviewers are trained to introduce the CASI section of the questionnaire and encourage respondents to participate –even if they haven't experienced a crime. At this point interviewers are asked to enter a response to Q163 which asks whether the respondent is happy to continue or not: 1. happy to continue (computer is passed to the respondent) 2. hesitant (respondent is hesitant to continue but interviewers are trained to reassure them and to help them get started) 3. refused totally (response option only used when the respondent completely refuses and the interviewer can't persuade them to continue with the CASI section). If an interviewer is not skilled at handling respondents' concerns if hesitant and encouraging participation – even if the respondent hasn't experienced a crime – respondents can drop out at this point of the questionnaire and hence the number of CASI victim forms could fall.	Refusal rates at the individual interviewer level were closely monitored and support was provided to any interviewer who appeared to be struggling to encourage people to take part. The consent rates below indicate that this has not been an issue for this iteration of the survey.

Quality measure	Description	Actual	Notes/Comments
CASI section skipped	Proportion of respondents who totally refused to complete the CASI section (ie Q163 = 3). Those that refused completely skipped the section, with the potential of lost victimisation data.	1.9%	The self-completion refusal rate in 2009 was 3.4% (Q163 = 3) and 6% in 2006. Lower refusal rates mean a better chance of collecting victimisation data which reduces non-response bias.
Reported self-completion	Proportion of respondents that reported in the audit telephone call that they completed a section by themselves using the computer. Independent check to ensure that respondents are given the opportunity to self- complete.	86.8%	These results are consistent with the results of Q409 data quality check which also suggest that 82.8% of respondents self-completed to some extent.
Recorded self-completion	Proportion of respondents that completed the CASI section with no, or very little, help from the interviewer (Q409 = 1 or 2).[32] Data collected from respondents that self- completed with little or no assistance from the interviewer is likely to be more honest and accurate than the data collected where the interviewer administered the questions.	82.8%	Given that the CASI section is being completed with no or little help from the interviewer in the majority of cases, it is likely that the responses recorded are true and accurate.

[32] More detail around this exit question can be found in Appendix C.

Respondents exit the interview prematurely	
Risk description	**Quality assurance processes**
Respondents could end the interview prematurely by not completing the CASI section and handing back the computer early.	Four surveys were abandoned: 1 survey at Q1, another at Q8, 1 at Q19.102 and 1 final survey at Q121 (VF2). One of these surveys contained incident data. For this survey 1 full victim form and 1 partial form were completed. A further 6 surveys were completed with non-Māori respondents in booster houses and were discarded. Interviewers were trained to provide support to the respondent when completing the CASI section. This included coaching them on the use of the laptop. Where a respondent decided after starting that they did not wish to use the laptop, the interviewer continued to administer the section.

Key exit questions

There are no quality measures associated with respondents exiting the interviews prematurely; however, a series of exit questions help us to monitor and understand why respondents either totally refused to participate in the CASI section (Q163 = 3) or didn't continue all the way through it.

Table 5.14: Back-coded main reason for not completing self-completion section

Question		Response	N	%
Q405	Please tell me your main reasons for not answering this (self-completion) section.	Too personal	30	22.9%
		Couldn't self-complete due to disability/literacy/IT literacy	29	22.1%
		Privacy concerns	16	12.2%
		Tired by this point/no time	14	10.7%
		Not interested	11	8.4%
		Too upsetting	10	7.6%
		Other	8	6.1%
		No reason given	7	5.3%
		No experience of this type of crime	6	4.6%
		Total	**131**	**100%**

Table 5.15: Other exit questions asked

Question		Response	N	%
Q409	Record how much you helped the respondent with the self-completion section.	Self completion done by respondent	5,109	75.0%
		One or two questions	530	7.8%
		More than one or two questions but less than half	123	1.8%
		More than half the questions but not all	151	2.2%
		All or nearly all of the questions	895	13.2%
		Total	**6,808**	**100%**
Question		**Response**	**N**	**%**
Q409.101	What type of assistance did you provide?	Read 1 or more questions to respondent	172	10.1%
		Read majority/all questions to respondent	461	27.1%
		Helped respondent enter one or more answers	340	20.0%
		Helped respondent enter majority/all of answers	888	52.3%
		Helped respondent move to the next screen	512	30.1%
		Helped respondent back up to previous screen	289	17.0%
		Answered questions about what the question meant	80	4.7%
		Other – Specify	23	1.4%
		Total	**2,765**	**100%**

Table 5.16: Fieldwork risks, quality assurance processes and outcomes

IT issues occur	
Risk description	**Quality assurance processes**
There are a number of IT issues that could impact the number of victim forms being submitted, it is up to interviewers to identify if and when these are happening (if they occur during the CASI section, identification will be more tricky) and report them to CBG management for resolution.	Where serious IT malfunction occurred in the field, and the interviewer was able to successfully reboot the laptop, they were able to re-launch the survey from the last question that was answered. This happened very rarely and there were no reports of surveys being abandoned because of this. There were no occurrences of serious IT failure or laptop theft that resulted in data being unrecoverable. Interviewers were trained to monitor respondents when completing the CASI section and were instructed to offer assistance if the respondent appeared to be stuck. There were no reports of any respondents starting the CASI section and not completing it due to IT issues.

Poor response rates	
Risk description	**Quality assurance processes**
A low response rate can lead to non-response bias, where the target population is not adequately represented in the survey. Non-response broadly comprises those people that refuse to take part in the survey and those that cannot be contacted. Ensuring that these people take part increases the accuracy and reliability of the results.	Continual response rate monitoring and reporting

Quality measure	Description	Result	Notes/Comments
Response rate	Response rate calculated as per previous iterations of the NZCASS.	81.0%	Main sample = 80.0%, booster sample = 84.8%. This compares with 71% main and 69% booster in 2009.
Respondent ethnicity distribution (core sample)	Response rates by ethnic group in the core sample. Response distribution in the core sample when analysed by ethnicity should very roughly match the ethnic diversity of the target population if interviewers are recruiting effectively.	European = 72.7% Māori = 12.9% Pacific = 4.7% Asian = 8.6% Other = 8.2%	See also Appendix E.

Quality measure	Description	Result	Notes/Comments
Male respondent proportion	Census data shows that males make up 47% of the adult population in New Zealand. Females live in smaller households on average than males, so will tend to predominate in the NZCASS because only 1 respondent is selected from each household. The unweighted proportion of males in the sample is thus expected to be lower than the census figure. Males have a different rate of victimisation to females. To ensure the survey is representative, male ratios are monitored at the individual interviewer level.	44.1%	Male ratio was 43.3% in the 2009 NZCASS. CBG achieved similar male rates in other population surveys. The ratios of male respondents in population surveys often differ from the national rate, due to features of the sample design or non-response. See also Appendix E.

Table 5.17: Other fieldwork quality measures

Quality measure	Description and purpose	Result	Notes/Comments
Enumerated houses in meshblock	Proportion of houses per area meshblock that were added into the sample by the interviewer. Results were analysed at an individual interviewer level to ensure that the in-field enumeration task was being completed correctly.	1.6%[33]	Rate demonstrated that enumeration was being completed as intended.
Adult phone number supplied in exit questions	Proportion of surveys with a phone number recorded in the exit questions for audit purposes. In order to conduct telephone audit calls, permission is requested from the respondent at Q407 and a number provided at Q408.	97.3%	
Phone number invalid or incorrect	Proportion of respondents with an incorrect or invalid phone number when contact was attempted by the auditing team. Phone numbers are used to conduct audit calls. A high level of accuracy is required when recording contact details to ensure all respondents have an opportunity to provide feedback via these calls.	6.5%	

[33] The 1.6% rate was calculated using the total selected addresses as the denominator.

Quality measure	Description and purpose	Result	Notes/Comments
Remembering completing survey	Proportion of respondents that remember completing the survey when asked in the telephone audit. Used to ensure that the survey was completed with the selected respondent recorded in the Sample Manager.	99.0%	Very occasionally a respondent will report that they did not remember the survey. This is more prevalent with elderly respondents or those that want to avoid answering any further questions. Where a respondent reports not remembering the survey, a GPS check is conducted to confirm that the surveyor was at the address for the duration of the survey.
Don't know/refused questions	Proportion of questions responded to with a 'Don't know' or 'Refused' response. Used as a general quality check to ensure item non-response rates are not too high for any individual interviewer.	2.2%	

Interview falsifying surveys	
Risk description	**Quality assurance processes**
If interviewers falsify surveys then the integrity of the data is compromised.	No evidence of survey falsification was detected in 2014.

Electronic audits

Electronic audits of data such as interview durations and question timings were also carried out; that is, survey paradata[34] was analysed. In particular, the electronic audits related to timings of interviews overall, and timings of sections of questions within the questionnaire. This data was analysed to check for outliers and anomalies that suggested problematic interviewer or questionnaire performance.

Individual interviewer performance was analysed with respect to interview durations, timing for specific questions, timing for groups of questions, and any questions or interviews which appeared to be entered or conducted out of hours (between 10:00pm and 8:00am).

[34] Survey paradata is information about the process of survey data collection. For further information about current developments with respect to survey paradata, please see O'Reilly (2009).

Checks of interview data

CBG conducted a number of ongoing checks of interview data throughout the fieldwork period and appropriate action was taken if any anomalies were discovered. Most of these checks were carried out on a weekly basis.

- Checks ensured that each laptop's date and time settings were correct by examining this data within each interview record.
- Checks were carried out for interview completeness, to ensure the last question in the demographics section had been answered in all interviews. Incomplete interviews were not included in the dataset.
- Checks were made to detect interviews with very short interview durations. CBG defined this as a questionnaire duration less than 10 minutes.[35] There were 55 interviews (0.8% of interviews) which had legitimate questionnaire durations of less than 10 minutes. As expected, none of these interviews[36] contained any CAPI or CASI victim form data.
- Checks were also made to detect interviews with unusually long interview durations. CBG defined this as questionnaire durations greater than 120 minutes.[37] There were 10 interviews[38] which had legitimate questionnaire durations longer than 120 minutes.
- Checks were undertaken for interviews which were in the Māori booster sample, but where the respondent had not selected Māori as one of their ethnic groups, and therefore the interview had been terminated near the beginning of the demographic questions. In total, during the interviewing period as a whole, 8 interviews were deleted from the dataset for this reason.

[35] This is the same as an interview duration less than 20 minutes.
[36] These 55 interviews were conducted by 16 different interviewers.
[37] This is the same as an interview duration greater than 130 minutes.
[38] These interviews were conducted by 3 different interviewers.

6. Fieldwork statistics

Introduction

This chapter provides detail about response rates and other key fieldwork statistics used as part of the NZCASS. Fieldwork statistics provide:

- measurement and monitoring information for research/fieldwork management
- useful information for planning future research
- an indication of issues or biases that may be present in the data and need to be noted or addressed.

Table 6.1: Summary of key fieldwork statistics by sample

	Main		Māori booster		Overall	
	2009	**2014**	**2009**	**2014**	**2009**	**2014**
Dwellings visited	6,934	7,990	14,008	3,464	20,942	11,454
Estimated eligible	6,809	6,540	1,893	2,015	8,703	8,574
Targeted number of interviews	4,030	4,800	1,409	1,660	5,439	6,460
Number of interviews achieved	4,809	5,235	1,297	1,708	6,106	6,943
Interview yield from dwellings visited	69%	66%	9%	49%	29%	61%
% of target completed (interviews achieved/target)	119%	109%	92%	103%	112%	107%
% of total sample	79%	75%	21%	25%	100%	100%

Response rates

Overview

Table 6.2: Unweighted response rates for the NZCASS since 2006

Year	Main	Māori booster	Overall	Notes
2014	80%	85%	81%	• Contact outcomes and response rate calculation were kept largely consistent with 2009. • 'Not visited' and 'Not a dwelling/Empty section' codes were added in 2014.
2009	71%	69%	70%	• Contact outcomes in 2009 for 'Unavailable' and 'Not available' are not directly comparable with the 'Unavailable' code recorded in 2006. • Similarly, codes for respondents who were not able to be interviewed were not directly comparable. In 2009, 'Language' and 'Incapacitated (infirm/hospitalised)' were used, whereas a single 'Respondent not interviewable' code was used in previous surveys.
2006	59%	56%	58%[39]	

Response rate changes

In order to reduce non-response and sampling bias, project teams in each iteration of the NZCASS have attempted to maximise response rates. Between 2006 and 2009, and again between 2009 and 2014, the proportional increase in response rates was relatively similar.

The increased response rate is likely to have reduced any non-response bias. The survey weights also combat non-response bias, primarily in the non-response adjustment and raking/post-stratification stages. These weighting stages are described in Chapter 10, along with the rest of the weighting process. One way to measure how much effect these weighting stages had is to compare the sample profiles after applying the initial inverse probability weights with the final weighted profile (these profiles are shown in Table E1 in Appendix E). On average, the figures changed by 0.8 percentage points in 2014. This is very similar to the average change of 0.9 percentage points in 2009, suggesting that the weights have had a similar overall effect in the last two iterations of NZCASS. In contrast, the average change in the 2006 profiles was 1.7 percentage points, roughly twice the size, indicating that the weights had to work harder in NZCASS 2006. This may well be due to the lower response rate in 2006.

While this analysis suggests there may have been greater underlying sample imbalances due to non-response in 2006 than in 2009 and 2014, these imbalances have been corrected by weighting. It is unknown whether substantive non-response bias remains after weighting, but it seems reasonable to suspect from analogy with the effect of weighting that NZCASS 2006 results might be more affected by any remaining bias than those from 2009 and 2014.

Appendix A also presents some demographic information (NZDep2013, 2013 Police crime rate and gender) for the households and respondents who declined to participate in the 2014 NZCASS.

[39] Not stated in the 2006 Technical Report.

Maximising response rates

To maximise the response rate in 2014, the following new procedures were implemented:

- the Māori booster sampling process was changed
- a pre-survey letter and pamphlet was sent to households prior to interviewer calling
- interviewer performance was monitored throughout the project with additional training and support being provided.

These procedures were in addition to that used during 2009 fieldwork to maximise response rates:

- a high maximum number of calls (10) to each dwelling (household) was used
- these (up to 10) calls were spread on different days, and at different times of the day
- there was a minimum of 5 visits to each meshblock
- using well-designed publicity and promotional materials in a variety of languages – in particular, the design and use of an information brochure in a question and answer format potential respondents could request an interviewer of the same gender or ethnicity as themselves, and make/change appointment times
- 0800 numbers for the Ministry of Justice, CBG and the Victims of Crime information line were prominently displayed on the brochure and letter, and the Victims of Crime website (www.victimsinfo.govt.nz) was also shown on the brochures
- respondents were informed about where and when they would be able to find the survey results
- promotion of the survey on the ministry's website was in place to increase awareness of the survey and provide evidence of authenticity

Contact outcomes

The same contact outcomes were used in 2014 as in 2009 so that response rates were as comparable as possible. Due to the 2014 change in sampling process, 2 new codes were added to this list of outcomes: 'Not visited' (NV) and 'Not a dwelling/Empty section' (NDE).

Interviewers recorded the outcome of the final call to each sampled dwelling as a code in Sample Manager. These outcome codes were then used in the response rate calculations. Please note that these were the final outcomes, as interviewers could call at a selected dwelling up to a maximum of 10 times.

Table 6.3: Contact outcomes, associated codes and categories

No.	Contact outcome	Code	Category
1	Interview	I	A
2	Not eligible	NE	B
3	Unavailable**	U	B
4	No reply	NR	C
5	Access denied/no access	AD	C
6	Household refusal	HR	D or C
7	Respondent refusal	RR	D
8	Not available**	NA	D
9	Appointment	APT	D
10	Language††	L	D
11	Incapacitated (infirm/hospitalised)	INC	D
12	Partial	P	D
13	Other	OTH	D
	Dwellings visited†		
14	Not visited	NV	C
	Estimated eligibles		
	Response rate (%)		
	Vacant*	V	*
	Not a dwelling/Empty section*	NDE	*

† 'Dwellings visited' was the sum of the 13 contact outcomes listed above. These were the occupied dwellings; the unoccupied dwellings (vacant dwellings) were listed separately.

†† This referred to English language difficulties; that is, household members could not understand the interviewer or any of the printed brochures.

* These contact outcomes (V and NDE) were not included in either the response rate calculation or the calculation of (occupied) dwellings visited, but has been included in this table for completeness. Note also that the 'out of frame' outcome was also excluded from the response rate calculations. There were 14 'out of frame' outcomes (12 from the main sample and 2 from the Māori booster).

** The difference between the 'Unavailable' and 'Not available' outcomes is that 'Unavailable ' referred to usual residents who were living away from the household for the duration of the survey, whereas 'Not available' referred to selected usual residents who were not available for the interview at the time of call by the interviewer.

Table 6.4: Summary of 2014 contact outcomes by sample

No.	Contact outcome	Main sample	Māori booster sample	Overall sample
1	Interview	5,235	1,708	6,943
2	Not eligible	0	1,061	1,061
3	Unavailable	260	60	320
4	No reply	328	115	443
5	Access denied/no access	200	65	265
6	Household refusal	445	143	588
7	Respondent refusal	117	45	162
8	Not available	62	19	81
9	Appointment	0	0	0
10	Language	45	1	46
11	Incapacitated (infirm/hospitalised)	103	36	139
12	Partial	3	0	3
13	Other	21	5	26
14	Not visited	4	2	6
	Dwellings visited	7,990	3,464	11,454
	Estimated eligibles	6,541	2,015	8,574
	Response rate (%)	80.0	84.8	81.0
	Vacant	706	149	855
	Not a dwelling/Empty section	465	57	522

Comparison of contact outcomes between 2009 and 2014

Table 6.5 provides a comparison of contact outcomes as a proportion of dwellings visited. This shows that:

- the number of interviews achieved as a proportion of the number of dwellings visited increased in 2014. Likewise the proportion of ineligible contacts decreased notably. These changes are likely to be in part the result of changes made to the Māori booster sampling process.
- the proportion of respondent refusals decreased, particularly as part of the main sample. There could be any number of reasons for this; however, the main 3 contributing factors are thought to be:
  - the interviewers selected as part of the NZCASS field team all had experience working on other large surveys with CBG. No new interviewers were recruited for the project.
  - interviewer performance was closely monitored throughout fieldwork, with additional support being provided where appropriate in order to reach targets
  - a pre-survey letter to the household was sent out in 2014 to help engage and prepare respondents for the survey.
- all other contact outcomes remained relatively stable when looked at as a proportion of all dwellings visited. This indicates that contact outcome coding was consistent with 2009.

Table 6.5: Contact outcomes by sample type, 2009 and 2014

		Main sample		Māori booster sample		Overall sample	
No.	**Contact outcome**	**2009**	**2014**	**2009**	**2014**	**2009**	**2014**
1	Interview	69%	65%	9%	49%	29%	61%
2	Not eligible	0%	0%	82%	31%	55%	9%
3	Unavailable	1%	3%	0%	2%	1%	3%
4	No reply	3%	4%	2%	3%	2%	4%
5	Access denied/no access	1%	3%	1%	2%	1%	2%
6	Household refusal	8%	6%	2%	4%	4%	5%
7	Respondent refusal	11%	1%	2%	1%	5%	1%
8	Not available	3%	1%	1%	1%	2%	1%
9	Appointment	0%	0%	0%	0%	0%	0%
10	Language	1%	1%	0%	0%	0%	0%
11	Incapacitated (infirm/hospitalised)	1%	1%	0%	0%	0%	0%
12	Partial	0%	0%	0%	0%	0%	0%
13	Other	1%	0%	0%	0%	0%	0%
14	Not visited	–	0%	–	0%	–	0%
	Dwellings visited	6,934	7,990	14,008	3,464	20,942	11,454
	Estimated eligibles	98%	82%	14%	58%	42%	75%
	Response rate (%)	70.6	80.0	68.5	84.8	70.2	81.0
	Vacant	7%	9%	5%	4%	6%	7%[40]
	Not a dwelling/Empty section	–	6%	–	2%	–	5%

[40] The rate of unoccupied dwellings according to the 2013 Census is 10.6%, up from 9.7% in 2006.

Response rate calculations

The response rate calculations used the outcome of the final call to each sampled dwelling that interviewers recorded. These outcomes were allocated to categories in the following manner for each of the PSUs in the sample: i = 1 to 1000.

Table 6.6: Contact outcomes and categories

Category	Outcomes
Interviews (a_i)	• Interviews (I)
Not eligible (b_i)	• Not eligible (NE) • Unavailable (U)*
Eligibility not established (c_i)	• No reply (NR) • Access denied/no access (AD) • Not visited • Household refusal (HR) in Māori booster sample*
Eligible non-response (d_i)	• Respondent refusal (RR) • Not available (NA) • Appointment (APT) • Language (L) • Incapacitated (INC) • Partial (P) • Other (OTH) • Household refusal (HR) in main sample*

* For main sample dwellings this outcome was included in the 'Eligible non-response' (d_i) category, for Māori booster sample dwellings this outcome was included in the 'Eligibility not established' (c_i) category.

An estimate of the eligible households within the PSU was calculated:

$$a_i + d_i \frac{c_i \times (a_i + d_i)}{(a_i + b_i + d_i)}$$

The response rate was the number of interviews achieved divided by the estimated eligible households, as shown below. This was the formula for calculating the response rate for each of the main (core) and Māori booster (screened) sample components within each PSU (meshblock).

$$\frac{a_i}{a_i + d_i + \frac{c_i \times (a_i + d_i)}{(a_i + b_i + d_i)}}$$

This reduced, or simplified, to the following:

$$\frac{a_i \times (a_i + b_i + \mathrm{d_i})}{(a_i + d_i)(a_i + b_i + c_i + d_i)}$$

The response rate for a group of PSUs was the average of the response rate for the individual PSUs, weighted by the estimated eligible households within each.

Response rate progress during fieldwork (2014)

As part of the 2014 NZCASS, an online dashboard was set up by the fieldwork provider so that fieldwork statistics could be monitored in real time. In addition to this, the ministry received a monthly fieldwork report which summarised the fieldwork statistics for the end of the relevant month. Table 6.7 provides a breakdown of (cumulative) progress and response rates provided in the monthly reports throughout the 2014 fieldwork period.

Table 6.7: Response rate progress by fieldwork month

No.	Reporting date	Number of completed meshblocks	Number of completed interviews	Main sample response rate (%)	Māori booster sample response rate (%)	Overall sample response rate (%)
1	9 March 2014	26	1025	25.6%	30.4%	26.7%
2	2 April 2014	99	2,245	39.4%	45.9%	40.9%
3	5 May 2014	202	3,959	54.3%	53.3%	53.8%
4	1 June 2014	481	5,183	56.8%	56.5%	56.6%
5	1 July 2014	823	6,832	77.5%	82.3%	78.5%
6	7 July 2014	1,000	6,943	80.0%	84.8%	81.0%

Response rates by demographic and geographic factors

Tables 6.8–6.13 show response rates broken down by various factors.

Table 6.8: Response rates by region

Region number	Region	Number of interviews	Number of meshblocks (PSUs)	Overall sample response rate (%)
01	Northland	317	37	87
02	Auckland	1,880	304	75
03	Waikato	812	100	81
04	Bay of Plenty	585	64	84
05	Gisborne	156	11	90
06	Hawke's Bay	344	39	89
07	Taranaki	199	26	88
08	Manawatu – Wanganui	377	55	83
09	Wellington	739	114	83
16	Tasman	108	14	93
17	Nelson	94	12	88
18	Marlborough	91	11	99
12	West Coast	45	8	88
13	Canterbury	788	129	84
14	Otago	279	53	77
15	Southland	129	23	72
	Total	**6,943**	**1,000**	**81%**

Table 6.9: Response rates by meshblock deprivation[41]

Level of area deprivation (NZDep2013)	Number of interviews	Number of meshblocks (PSUs)	Overall sample response rate (%)
1 (lowest)	1,162	194	81
2	1,246	197	81
3	1,446	221	80
4	1,438	198	82
5 (highest)	1,633	187	81
NA	18	3	75
Total	**6,943**	**1,000**	**81%**

Table 6.10: Response rates by Police recorded crime groups

Crime Rate (2013)[42]	Number of interviews	Number of meshblocks (PSUs)	Overall sample response rate (%)
Low	2,336	381	81.8
Medium	1,830	270	79.8
High	2,777	349	81.1
Total	**6,943**	**1,000**	**81.0%**

[41] Three meshblocks in 2014 did not have a deprivation score.

[42] Crime rate groups have been derived for each meshblock from Police recorded crime data. The Police data relates to the number of incidents recorded as crimes (in -scope in NZCASS) that occurred in 2013. This produces frequency counts per meshblock, which are grouped into three evenly sized groups weighted by the 2013 estimated resident population 2013.

Interview counts by age, ethnicity and sex

Table 6.11: Ethnicity by total response

				Ethnicity									
				European		Māori		Pacific		Asian		Other	
Age group	Total	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
15–19 years	380	172	208	110	118	77	102	21	23	14	17	6	4
20–29 years	924	393	531	242	320	158	253	28	41	50	49	9	4
30–39 years	1,061	457	604	294	378	159	249	26	28	55	62	13	14
40–49 years	1,297	591	706	414	487	213	290	28	29	43	44	10	7
50–59 years	1,174	532	642	376	448	169	227	15	28	44	36	6	9
60–64 years	541	228	313	180	238	53	94	10	11	7	13	3	2
65 years and over	1,560	683	877	581	741	148	191	15	11	14	17	2	5
Refused	6	3	3	2	3	0	1	0	0	1	0	0	0
Total	**6,943**	**3,059**	**3,884**	**2,199**	**2,733**	**977**	**1,407**	**143**	**171**	**228**	**238**	**49**	**45**

Table 6.12: Gender by total response

				Ethnicity									
				European		Māori		Pacific		Asian		Other	
Sample	Total	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Main	5,235	2,374	2,861	1,894	2,274	292	384	118	126	221	233	47	43
Māori booster	1,708	685	1,023	305	459	685	1,023	25	45	7	5	2	2
Total	**6,943**	**3,059**	**3,884**	**2,199**	**2,733**	**977**	**1,407**	**143**	**171**	**228**	**238**	**49**	**45**

Table 6.13: Interviews completed by month

Month	Number of days	Number of interviews	%
10–28 February 2014	19	629	9.1
1–31 March 2014	31	1,519	21.9
1–30 April 2014	30	1,577	22.7
1–31 May 2014	31	1,453	20.9
1–30 June 2014	30	1,640	23.6
1–6 July 2014	6	125	1.8
Total	**147**	**6,943**	**100.0**

Completion of the questionnaire

In the NZCASS there are 4 key questionnaire completion milestones:

1. a respondent has completed the interviewer-administered (CAPI) section
2. a respondent has completed up to and including the demographics (Q163)
3. a respondent has completed the self-completion (CASI) section
4. a respondent has completed the exit questions.

Within some of these completion milestones, there are different ways that 'completion' can be measured.

As the standard, an NZCASS questionnaire was considered 'complete' for the purpose of inclusion in the final dataset if a respondent had completed up to and including Q163 – that is, completed the CAPI section up to and including the demographic questions.

Table 6.14 provides statistics for each measure used to indicate whether or not a questionnaire milestone has been completed.

Interview duration

The total average interview duration is made up of:

- the time it takes the interviewer to engage the respondent and 'get settled' ready to start the interview and disengage at the end of the interview
- administering the CAPI section of the questionnaire
- the time it takes for the respondent to answer the self-completion component of the questionnaire (CASI section)
- administering the exit questions and closing the interview.

Table 6.14: Average interview duration by number of victim forms completed, by year

Average interview duration (minutes)	2006	2009	2014
Total	**52**	**48.8**	**41.0**
No CAPI victim forms	44	41.4	36.5
One CAPI victim form	56	52.5	47.2
Two CAPI victim forms	66	61.8	56.3
Three CAPI victim forms	75	73.1	68.7
One or more CAPI victim forms	53	63.1	51.6

The average interview durations noted above include a 10-minute estimate to engage the respondent and exit from the interview. This estimate was first introduced in 2006 and then continued in 2009 because enumeration and approach processes were paper based, rather than electronic as in 2014. To estimate equivalent durations, the same 10-minute buffer has been applied in 2014; however, it is highly recommended that this is dropped as part of the process in future years and if technology allows, the engagement and exit timings are monitored (rather than estimated).

7. Offence coding

Introduction

Offence coding is an important part of the NZCASS. Offence codes are based on the responses provided in the CAPI and CASI victim forms, including a short description in the respondent's own words (except for sexual offences). Coding also draws to some extent on other questions throughout the questionnaire.

Offence coding activities were the least well-documented and structured part of the research process in previous iterations of the NZCASS. In 2014, a number of systems and processes were developed and implemented in order to:

- improve the transparency around offence coding practice
- improve the consistency of coding both between iterations of the NZCASS and between coders
- update offence coding resources so that they were easy to use and understand (eg the Offence Coding Manual)
- facilitate 'in-field' monitoring and reporting of coding progress
- facilitate 'in-field' monitoring and reporting of quality assurance activities as well as keep a record of all coding decisions and discussions undertaken
- facilitate recording of activity time to assist with the planning of future iterations.

It should be noted that while the management and IT systems used to conduct coding activities have changed notably in 2014, the project team took great care to ensure that the methods and rules used to code offences remained consistent with 2009 and 2006.

Offence coding resources

A number of resources were provided to 2014 coders both as part of their training and for use during coding activities.

Table 7.1: Offence coding resources

Resource	Description
Offence Coding Workbook	The workbook was a new resource in 2014 and provided: • contextual information about the survey • guidelines on work practice • a user's guide to the coding system/interface.
Offence Coding Manual	The purpose of the manual was to: • explain the principles of offence coding as part of the NZCASS • document coding practices and procedures. The Offence Coding Manual was heavily reformatted in 2014 so that it was easier to use (find information), understand and update. Some additional information was also included in 2014 such as a list of incompatible codes and a list of commonly used terms and definitions.
Questionnaire	A copy of the final questionnaire that was being used as part of 2014 fieldwork.
Offence code summary sheet	A 1-page 'lift out' that listed all offence codes (Appendix A of the Offence Coding Manual).
Crimes Act	Link to the Crimes Act so that coders could look up or check details should they need to.
Training presentations	Coders also had access to a range of online presentations delivered as part of the training.
Experts/Supervisors	Experts from both Victoria University and the New Zealand Police were available to assist coders as and when required throughout the coding process.

Coders and training

Coders

Because offence coding as part of the NZCASS requires a foundation in legal theory, the coders hired needed to:

- be fourth year honours students (law)
- have completed the criminal law module and legal reasoning/research modules at a B grade or above
- be able to give evidence of IT literacy
- have a high attention to detail.

Two coders were hired by CBG as part of the NZCASS pilot study and continued as part of the main study as well. In total, 6 coders worked on the NZCASS main study along with a CBG coding supervisor and 2 supporting auditors/experts (1 from the Victoria University Law School and 1 from the New Zealand Police).

Training as part of the offence coding process took place in a number of stages (Table 7.2).

Table 7.2: Training undertaken for offence coding

Resource	Description
Provision of workbook	Reading/self-learning.
Online webinar	• 1.5–2 hours • overview, orientation and demonstration.
Provision of manual	Reading/self-learning.
Individual practice	Navigating the system/interface.
Online assessment	A pass of 100% was required before starting live coding.
One day face-to-face training	Conducted at the Ministry of Justice in Wellington and covered the following: • NZCASS overview, past surveys and the importance of the NZCASS to the sector and how coding is an important part of the research process • Police recording standards and comparison with legal theory • a recap of the orientation webinar • offence coding, double coding, out-of-scope rules, key considerations, process and examples from previous iterations (2006 and 2009) • individual/group practice • group discussion on coding decisions and issues.
Individual practice time	• 5–7 days • using 2006 and 2009 records.
Observed assessment	Coders were subject to an online, observed assessment by the Victoria University expert with the assistance of the coding supervisor. A minimum of 6 CAPI/CASI victim forms were selected for coding during the assessment, which aimed to ensure that the coder could demonstrate the following competencies: • assign standard offence codes with a high degree of accuracy to offence data collected from the 2014 NZCASS pilot for both CAPI and CASI victim forms • coding decisions are based on a review of all the detail provided for each offence, including all forms for that victim • knowledge of when to submit a record as certain and when to submit as not certain and enter sensible, succinct and understandable comments as appropriate • refer back to the Offence Coding Manual before applying a code, in particular where an offence is borderline • ability to code with a high degree of accuracy common ambiguous and/or difficult offence scenarios, including double coding and 80s codes. Coders were able to commence live coding on real data once the assessors were satisfied that all of the above competencies had been met.

Coding practice and processes

The following section provides an overview of the offence coding and quality assurance process undertaken in 2014. Details of how offences were coded are provided in the NZCASS Offence Coding Manual.[43]

Overview

1. One of the NZCASS research objectives requires comparison with levels of reported crime. As such, it's important that offence coding for NZCASS mirrors Police recording practice as closely as possible.
2. An exact match with Police recording practice is unlikely given that:
   a. different Police officers may make different judgements when deciding:
      i. whether to record an incident as an offence
      ii. which category it should be placed in.
   b. Police continuously review and refine recording rules, which means some practice change, has occurred between surveys.
3. As a general principle, offences in the NZCASS are coded:
   a. in accordance with current legal theory
   b. in line with current Police recording procedures.
4. In most circumstances these 2 requirements will be met and there will be no conflict (ie Police recording practice will be in line with the legal theory and definitions).

Offence codes

Table 7.3 lists the offence codes collected in the NZCASS. These offence codes are unchanged between 2006, 2009 and 2014.

Table 7.3: Offence codes collected in the NZCASS

Offence label	Offence code	Not counted	Weight (H = Household P = Person
Sexual violation of women	01		P
Sexual violation of men	02		P
Incest	03	✓	
Indecent assault	04		P
Indecent exposure	05	✓	
Grievous assaults	06		P
Other assaults	07		P
Abduction/kidnapping	08		P
Robbery	09		P

[43] The NZCASS Offence Coding Manual is available from the ministry on request.

Offence label	Offence code	Not counted	Weight (H = Household P = Person
Theft from person	10		P
Burglary (old definition)	11		H
Burglary (new definition)	41		H
Theft from inside home (right to be there)	12		H
Theft from outside the home (over $10)	13		H
Taking/conversion motor vehicle	14		H
Unlawful interference/getting into motor vehicle	15		H
Theft from motor vehicles	16		H
Taking/conversion/unlawful interference with bicycle	17		H
General theft of personal property	18		H
Arson	19		H
Wilful damage to household property	27		H
Wilful damage to personal property	28		P
Threatening to kill or assault/threatening behaviour	21		P
Threatening to damage property	29		P
Extortion/blackmail	22	✓	
Unlawfully in building (no clear intention to commit offence)	23	✓	
Peeping Toms, lurking etc	24	✓	
Fraud	25	✓	
Damage to motor vehicles	26		H
In scope, but not able to tell which offence	85	✓	
Not an offence	86	✓	
Offence not in scope	87	✓	

Coding period

In 2014, offence coding took place in 'real time' while interviewing was taking place. This was possible due to new IT and management systems put in place. In previous iterations of the NZCASS, coding took place in batches once interviews had been completed.

The main study coding period officially ran between 2 March 2014 and 13 July 2014. The coding period finished 14 days after fieldwork once final coding and quality assurance processes were undertaken.

Coding interface/system

In 2006 and 2009, survey information was accessed and coded using a Microsoft Excel spreadsheet. Comments would be made on each form/record, which were then flagged for review or discussion with the wider team or coding supervisor.

In 2014, a coding interface/system was developed to help improve the quality of the data and assist with coding management and quality assurance processes. The new interface was an online, web-based system[44] designed by CBG. This system allowed coders to work remotely and around their other work and study commitments.

The advantages of the 2014 interface/system include:

- ease of navigation and ability to view all the information on 1 page for each respondent
- ease of moving between forms, an important consideration in ensuring all forms are reviewed before a final coding decision is made, to ensure that identical incidents are not coded more than once and to easily see any patterns of victimisation
- no delay in the survey data collected by the interviewer being made available to the coder – new records were loaded on a daily basis as interviewing progressed, thus reducing time pressure on the coding activity
- easier analysis and quarantine of coding decisions
- ability to limit access, tailor separate views for specific coders or auditors (eg only auditors could write in the auditor comments box and each coder sees their own individual list of records to be re-coded upon log in)
- instant reports in real time of the number of records submitted as certain/uncertain, outstanding for audit, re-coding, by whom and when
- shared comments and data in real time, preventing the need for spreadsheets going backwards and forwards between the coders and the expert team member – this method improved communication greatly and allowed for more efficient time management.

Screenshots of the 2014 coding interface/system have been provided in Appendix D.

[44] The system used the 'FileMaker' database platform.

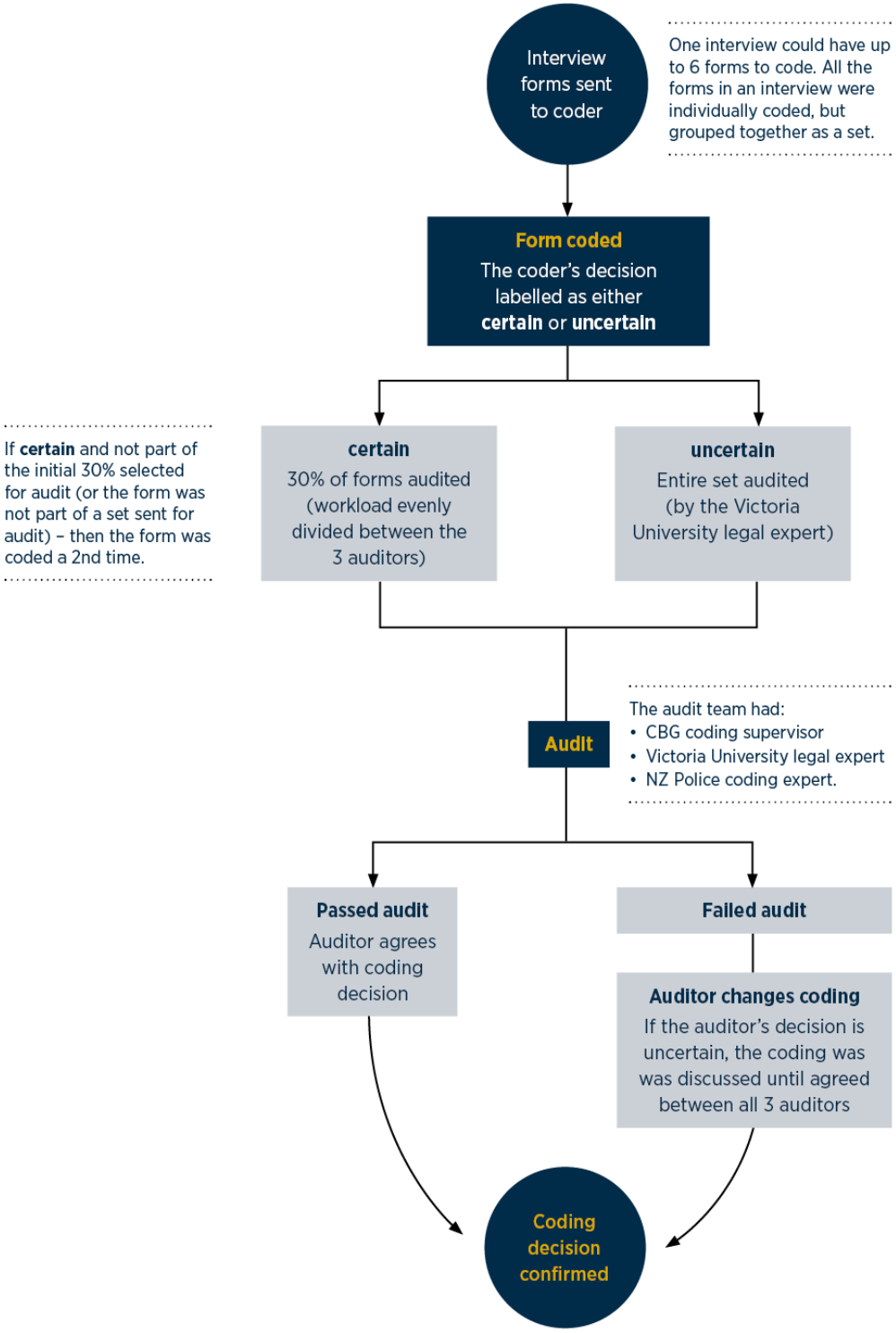
Coding quality assurance

To ensure that coding decisions were correct, a number of quality assurance steps were put in place in 2014 (Table 7.4).

Table 7.4: Offence coding quality assurance process

Step		Description
1	Forms sent to coder	One interview could have up to 6 forms to code. All the forms in an interview were individually coded, but grouped together as a set.
2	Certain vs Uncertain	Each coding decision needed to be marked as either 'Certain' or 'Uncertain' by the coder.
3	Certain codes	Ten percent of 'Certain' decisions were randomly assigned to each of the coding supervisors/auditors (30% of 'Certain' decisions in total). A balance of views, skills and experience was provided by 3 auditors: 1. CBG coding supervisor 2. Victoria University legal expert 3. New Zealand Police coding expert.
4	Uncertain codes	All decisions where the coder was uncertain of the offence code assigned were audited by the Victoria University legal expert.
5	Audit (pass/fail)	Where the decision passed the audit, the offence code was confirmed. Where the decision failed the audit: • the coding decision was changed by the auditor • comments were provided around why the code was changed and the rationale for this change • the record was sent back to the coder for review.
6	Auditor agreement	All 3 auditors would discuss the record and come to an agreed decision where: • A 'certain' decision was changed during the audit process • Where an auditor was uncertain about the correct offence code to assign
7	Coder check	Where a pattern of failed 'Certain' decisions emerged, a wider review of the coder's work would be triggered to assess if there were any 'patterns of concern' that needed to be addressed. This was done to ensure that all coders were working correctly and to the required standard. It should be noted that during the coding period, only 1 instance was flagged where a coder needed further coaching on how to assign primary vs secondary offence codes.
8	Secondary coding	As an added level of quality assurance, secondary coding was triggered in some cases. This means that in some cases, records were put back into the system so that they could be coded a second time by a different coder. This was done so that mismatches could be identified and further auditing could take place.

Figure 7.1: Offence coding flow process



Quality assurance statistics

Quality assurance statistics were reported to the ministry monthly throughout the fieldwork period. Table 7.5 shows the final number of coding decisions audited along with pass rates for 2014.

Table 7.5: Offence coding quality assurance statistics

	Initial coding process	Secondary coding process	Notes
Total number of records coded	3,755	772 (21%)	The secondary coding process figure refers to the number of records that were coded by 2 different coders as another level of quality assurance.
Number of mismatches	N/A	35 (4.5%)	Number of mismatches between the codes assigned by the initial and secondary coder. All mismatched codes were reviewed by auditors.
Number of 'Certain' records	2,818 (75%)	571 (74%)	
Number of 'Certain' records selected for audit	873 (31%)	229 (40%)	Ten percent of 'Certain' records were randomly selected for audit by each of the 3 auditors.
Audit pass	803 (92%)	197 (86%)	Secondary coding process: Number of records that were coded for a second time, did not have a 'mismatch' flag, were selected for audit and passed.
Audit fail	70 (8%)	32 (14%)	Secondary coding process: Number of records that were coded for a second time, did not have a 'mismatch' flag, were selected for audit and passed.
Number of 'Uncertain' records	937 (25%)	201 (26%)	
Number of 'Uncertain' records selected for audit	937 (100%)	201 (100%)	
Initial code correct	615 (66%)	150 (75%)	
Initial code incorrect	322 (34%)	51 (25%)	

Coding process timings

Table 7.6: Offence coding process timings

	Pilot study	Main study
Average number of forms coded per hour	25 forms (1 form on average every 2.5 minutes)	30 forms (1 form on average every 2 minutes)
Average number of forms audited per hour (quality assurance process)	20 forms	20 forms
Average time worked per coder per week	NA	1.3 hours
Average time worked per auditor (quality assurance expert) per week	NA	2 hours
Total coding time	NA	260 hours (13.7 hours for each week of fieldwork)

Offence coding statistics

Number of forms coded

Table 7.7: Total number of forms coded

Type of Form	2014
VF1	1,939
VF2	765
VF3	409
SC1	188
SC2	345
SC3	92
Total	**3,738**[45]

Distribution of offence codes

Table 7.8 looks at the distribution of primary and secondary offence codes obtained in CAPI and CASI for each year. While there have been some minor changes, the distribution of codes between years has remained relatively consistent indicating that coding practice has also remained relatively consistent between years.

Table 7.8: Distribution of offence codes, by year

Offence code	2006	2009	2014
1	0.9%	0.7%	0.8%
2	0.1%	0.1%	0.1%
4	2.1%	2.0%	1.5%
5	0.0%	0.0%	0.0%
6	1.1%	1.2%	0.8%
7	8.6%	8.0%	7.1%
8	0.0%	0.1%	0.0%
9	0.6%	0.9%	0.6%
10	0.8%	0.3%	0.5%
11	8.3%	6.8%	8.0%
12	4.0%	4.4%	3.9%
13	1.2%	1.7%	1.0%
14	3.4%	2.4%	2.6%

[45] Of these 3,738 coded victim forms, offences out of scope for NZCASS (eg 80s offence codes) and incomplete victim forms were excluded for further analysis. This resulted in 2,824 incidents in the final analysis dataset (discussed further in Chapter 8).

Offence code	2006	2009	2014
15	0.5%	0.4%	1.0%
16	9.2%	7.8%	8.2%
17	1.4%	1.8%	1.7%
18	3.5%	3.2%	2.6%
19	0.3%	0.3%	0.2%
21	7.6%	6.5%	7.8%
23	0.7%	0.3%	0.4%
24	0.1%	0.0%	0.1%
25	0.1%	0.5%	0.9%
26	8.1%	7.6%	7.3%
27	10.0%	9.4%	8.4%
28	1.5%	1.4%	1.1%
29	1.3%	1.3%	1.5%
41	10.8%	13.2%	15.1%
85	0.6%	1.0%	1.0%
86	7.5%	12.2%	11.3%
87	5.7%	4.9%	4.5%

Note: 0.0% is either rounded to 0 or nil.

Double coding

An incident that has different elements can have more than 1 offence code applied under certain situations. Multiple offences can be coded from both the victim forms and self-completion information. Table 7.9 presents the percentage of offences that were double coded in each of the 3 survey years.

Table 7.9: Number of double coded offences, by year

Survey year	Number of double coded offences	Total number of coded offences	Percentage of double coded offences
2006	292	4,573	6.4%
2009	501	5,493	9.1%
2014	332	3,738	8.9%

8. Data processing

Datasets

Each interviewer was required to upload encrypted survey data to CBG servers every day they were active in the field. The files consisted of all changes that had been made to the Sample Manager database residing on the interviewer's laptop since the last upload. For example, this could include new survey data, information on contact attempts or new household outcome coding.

Once received at CBG, the files were decrypted and checked before being processed into a SAS data warehouse. A number of datasets resided within the warehouse pertaining to survey data collected via the TSS questionnaire, exit questions (recorded directly into the Sample Manager) and other survey metrics recorded by the interviewer (eg respondent information and outcome coding).

The contents of each export file were analysed and directed to the relevant datasets ready for further formatting and cleaning. Data pertaining to the offence coding process was entered directly into a secure web interface which wrote directly to its own SAS dataset.

Once the survey data had been formatted and cleaned, several output datasets were created for delivery to the ministry (see Table 8.1). Final datasets were delivered to the ministry on 31 July 2014.

Table 8.1: Datasets delivered to the ministry by the fieldwork provider

Dataset	Description	Supplied format
Main	Contains all variables relating to the questionnaire; that is, responses captured in the attitudes and perceptions of crime section, CAPI screeners and victim forms, demographics section and CASI screeners and victim forms. The main dataset also includes survey duration data and derived ethnicity variables.	SAS dataset (.sas7bdat)
Offence coding	Contains offence codes assigned to all incidents recorded in the questionnaire along with information on the auditing process and outcome.	Excel spreadsheet (.xlsx)
Household outcomes	Contains information on the final contact outcomes of all selected addresses in the sample.	Excel spreadsheet (.xlsx)
Datamatching	Contains information collected as part of the datamatching consent process for those respondents that agreed to this part of the survey.	Excel spreadsheet (.xlsx)
Recontact	Contains information collected as part of the recontact consent process for those respondents that agreed to being contacted to take part in further research.	Excel spreadsheet (.xlsx)

Formatting

Questionnaire responses arrive from the field as raw survey files. Formatting of this raw data was performed to ensure that the supplied datasets were consistent with the questionnaire document and datasets provided in previous iterations of the survey. The following tasks were undertaken during the formatting stage:

1. Variables were renamed to match the question numbers used in the questionnaire document.
2. Unwanted variables were removed. These were usually 'dummy' variables that were included in the survey in order to achieve desired functionality and behaviour required (eg complex skip logic and consistency checks).
3. Question responses were re-coded to match the questionnaire document. Occasionally response options were assigned different numbers to the questionnaire document. For example, Q9 has response options ranging from 0–10 in the questionnaire document. As TSS cannot record a zero response, the survey file contained responses in the range 1–11. These responses were adjusted in the formatting stage to be consistent with the questionnaire document.
4. Variables were merged. There are many questions in the survey where the text of the individual question varies depending on responses to previous questions. For example, most of the victim form screener questions have 2 different versions, depending on whether any incidents had been reported earlier in the screener section. Each of these question variants required a separate variable to be created in the survey data file. As the responses to these different variants related to the same question, the different versions of the question were merged into a single variable.
5. Multiple response questions were converted into binary flag variables where every response in the answer framework was assigned 0 or 1 to indicate if the response had been selected.
6. Survey responses recorded in Sample Manager were merged into the main dataset. In 2014, the exit questions were administered in CBG's Sample Manager software. The responses to these questions needed to be combined with the responses recorded in the survey software.
7. Derived variables were added. For the main dataset, these were the duration and ethnicity code variables.
8. Variables were reordered to match the questionnaire document.

Data cleaning

During the interview process, respondents sometimes decide to go back to a previous question and change the response that was originally provided. Occasionally when the response is changed, the respondent may branch off to a different part of the survey as a result of this. By default, the response recorded on the old logic path persists in the data file.

For example, a respondent answers 'no' to Q3, the logic path takes them to Q4, which they provide an answer for. The respondent then decides to go back to Q3 and change their answer to 'yes', this action sets them on a different logic path which skips Q4 and takes them directly to Q5. The response recorded at Q4 will still be present in the survey file.

These unwanted responses, known as skip violations, were cleaned from the dataset prior to delivery to the ministry. Failure to remove these responses could have resulted in incorrect estimates being reported for the affected questions.

The cleaning process involved SAS code being written to mimic the skip logic of the survey. This code was run through the uncleaned dataset, thus removing all responses which appeared to violate the skip logic of the survey. Next, the uncleaned and cleaned datasets were compared and differences identified for investigation. Every response that was deleted from the uncleaned dataset was checked to ensure that it had been correctly removed (ie that a respondent had indeed gone back and changed an answer which caused them to proceed on a different logic branch). Checking of every violation ensured that no data was erroneously deleted.

The cleaning code underwent quality assurance processes by both CBG and ministry staff and was refined following delivery of both interim datasets to account for logic conditions that were not previously included. Uncleaned and cleaned datasets along with the cleaning code were supplied to the ministry[46].

[46] Within most survey software cleaning functionality is included and this process is not required. Due to the complexity of the NZCASS questionnaire and weaknesses in the survey software used by CBG, the implementation of this cleaning programme was necessary.

Data quality assurance

Two interim datasets were provided to the ministry on the completion of 200 and 1000 interviews. The purpose of these interim datasets was to check the quality of the data provided and, where necessary, resolve issues or strengthen quality assurance processes ahead of the final dataset delivery.

Prior to delivery, all datasets were subject to a number of checks developed by the ministry, Statistical Insights and CBG. The checks were completed by CBG using SAS with the results being provided to the ministry in a report accompanying the datasets. Table 8.2 summarises the main checks that were conducted on each of the datasets supplied. (Note: This is not an exhaustive list of all checks that took place.)

Table 8.2: Data from fieldwork provider quality assurance checks

Dataset	Checks undertaken
Main	• Sample sizes were as expected. • Question outliers were identified and investigated. • Inconsistencies within and between questions were identified and investigated. • Missing or unexpected values were identified and investigated. • Questionnaire sections were complete for the vast majority of respondents. • CAPI and CASI victim forms were complete for all selected incidents. • Question timings were recorded for all questions and question sections. • Refusal rates were at or below expected levels.
Offence coding	• Offence codes were assigned to all suitable incidents recorded in the CAPI and CASI victim forms. • Records were audited per the agreed algorithm. • Information on the outcome of the audit was recorded including details of any updated offence codes. • Where 2 offence codes were assigned, these were compatible and ordered correctly.
Household outcomes	• Final outcome codes were assigned to all selected dwellings. • Unique ID numbers were assigned to each household.
Datamatching	• Date of birth information provided for the purpose of datamatching was consistent with age group recorded in survey. • Date of birth provided was within a sensible range. • Surnames were provided for the vast majority of respondents that agreed to datamatching.
Recontact	• Contact details were recorded for all respondents who agreed to take part in future research. • Contact details were recorded for all respondents who requested to receive a copy of the summary survey findings.

Analysis datasets

Once data had been supplied and undergone quality assurance, the ministry project team converted the datasets into final 'analysis datasets'. This involved:

1. renaming the data items (variables) to useable SAS names as specified in the data item list, rather than being based on the question number
2. creating the 'incident' dataset by combining the victim form information collected in up to 6 victim forms per respondent, while excluding the offences out-of-scope for NZCASS and for incidents with incomplete victim forms.
3. merging on other data as required, including NZDep2013, region and urban classifications (further detail in Chapter 9)
4. deriving data items required for analysis, such as merging categories together or combining multiple data items (such as combining tenure and landlord type into a combined data item), and coding household composition (further detail in Chapter 9)
5. grouping offences together
6. general tidy up to ensure consistent code frames and specification of responding populations.

Quality assurance processes were undertaken for each of these steps.

Statistical Insights then derived the personal and household weights and replicate weights and undertook imputation for the survey, which was then merged on the analysis datasets.

Table 8.3 shows the 3 datasets constructed through this process and used for analysis.

Table 8.3: Analysis datasets

Analysis dataset	SAS name (2014)	Level (2014 number rows)	Description	Use
Main	MAIN14	Person level Rows = 6,943	Contains demographic, geographic, attitude and perceptions, plus the household and person weights	• attitude • perceptions • demographic • geographic
Incident	INCIDENT 14	Incident level Rows = 2,824	Contains all information on victim forms (up to 6 in total), plus the incident weights	• victim • experiences • victim needs • reporting to Police
Imputation	MIOFFEN CES14	Multiple imputed Person level × 100 Rows = 694,300	Contains the incidence and prevalence items for each offence and offence group, plus the household and person weights	• incidence rates • prevalence rates • total crime rates • victimisation distributions

The derivation of the analysis datasets was also retrospectively applied to the 2006 and 2009 surveys to ensure consistency and comparability between the survey years.

9. Classifications, coding and groupings

Introduction

This chapter provides detail about the classifications used to output the data and how offences were grouped together for analysis. The changes made from previous iterations of the NZCASS are as follows:

- all classifications were reviewed and where possible brought in line with Statistics NZ standard classifications
- how offences were grouped together was reviewed based on current stakeholder need, conceptual appropriateness and what is possible due to sample sizes for small offences.

Classifications

A statistical classification is a way to group a set of related categories in a meaningful, systematic and standard format. The value of statistical data is maximised when classified in a consistent way across data sources.

While the 2009 NZCASS used a number of standard classifications, such as the statistical standard for ethnicity, the 2014 project team conducted a full review of the classifications used on demographic and geographic measures to bring new and existing demographic and geographic measures in line with Statistics NZ standard classifications.

The decision rules applied during this review were:

1. In the first instance, align to Statistics NZ standard classification if appropriate.

   If the Statistics NZ standard classification was either (a) not appropriate due to NZCASS reporting or (b) there was not the sample size to report on each category in the standard classification, then:

2. Align to Statistics NZ standard classification as closely as possible, with variations required due to NZCASS reporting or to support sample sizes.

   If the standard is not an appropriate fit for NZCASS reporting or not possible due to the way the information is collected in the NZCASS questionnaire, then:

3. Continue with the classifications used in 2009 and 2006 if sensible.

Table 9.1 shows the final demographic and geographic classifications used for 2014 NZCASS reporting, along with the sample sizes for each category.[47]

[47] Residual categories not output (such as 'Don't know' and 'Refused') are not presented, hence the sample sizes for each data item may not sum to the total number of respondents.

Table 9.1: Sample sizes by demographic and geographic classifications

Data item	Categories	Sample size (2014)
Personal factors		
Sex	Male	3,059
	Female	3,884
Age	15–19 years	380
	20–29 years	924
	30–39 years	1,061
	40–49 years	1,297
	50–59 years	1,174
	60–64 years	541
	65 years and over	1,560
Ethnicity	European	4,932
	Māori	2,384
	Pacific peoples	314
	Asian	466
	Other ethnicity	94
Legally registered relationship status	Married/civil union (not separated)	2,985
	Separated/dissolved	995
	Widowed/surviving partner	602
	Never married and never in a civil union	2,339
Partnership status	Partnered	
	Legally registered	2,897
	Not legally registered	1,171
	Non-partnered	2,727
Economic factors		
Employment status	Employed	4,011
	Unemployed	244
	Not in the labour force	
	Retired	1,394
	Home or caring duties	481
	Studying	436
	Not actively seeking work/unable to work	334
Financial stress: Limited to buy item for $300	Not at all limited	3,087
	A little/quite limited	1,856
	Very limited/couldn't buy it	1,973
Financial stress: Can meet unexpected expense	Yes	5,321
	No	1,558

Data item	Categories	Sample size (2014)
NZ Deprivation Index	Quintile 1 (least deprived) Quintile 2	1,162
	Quintile 3	1,246
	Quintile 4	1,446
	Quintile 5 (most deprived)	1,438
		1,633
Personal income	$30,000 or less	2,968
	$30,001–$70,000	2,314
	$70,001 or more	909
	Don't know/Refused	752
Household income	$30,000 or less	1,280
	$30,001–$70,000	1,959
	$70,001–$100,000	922
	$100,001 or more	1,410
	Don't know/Refused	1,372
Household factors		
Household composition	One person household	1,518
	One parent with child(ren)	643
	One parent with child(ren) and other person(s)	198
	Couple only	1,838
	Couple with no children and other person(s)	130
	Couple with child(ren)	1,693
	Couple with child(ren) and other person(s)	168
	Multiple family household	177
	Other multi-person household	330
Tenure and landlord type	Owned (including with a mortgage)	4,561
	Rented – private	1,848
	Rented – government (local and central)	448
Geographic factors		
Urbanisation	Main urban area	4,778
	Secondary urban area	447
	Minor urban area	763
	Rural	955
Region	Auckland	1,880
	Wellington	739
	Rest of North Island	2,790
	Canterbury	788
	Rest of South Island	746
Total respondents		**6,943**

Demographic coding

The 2 demographic questions recoded as part of the NZCASS processing period were ethnicity and household composition. This section also describes how 'Other – Specify' options were handled.

Ethnicity

CBG coded the responses to the ethnicity question Q151 to the 2-digit Ethnicity New Zealand Standard Classification (2005).

The survey was pre-loaded with the Statistics NZ database of ethnicity classifications. The addition to the code frame was a separate code assigned to responses of 'New Zealander' and/or 'Kiwi' to code

62. If a respondent selects they are of 'Other' ethnicity at Q151 they are taken to a second screen where the 'Other' ethnic groups are recorded. As the interviewer starts to type into the text box, a list of matches from the database are displayed, and the correct ethnic group can be selected. This process provided CBG with a 2-digit ethnicity classification for ethnicity.[48]

The multiple response 2-digit codes were provided to the ministry and each respondent was assigned into ethnic groups as follows:

- European – 10, 11, 12, 62
- Māori – 21
- Pacific peoples – 30, 31, 32, 33, 34, 35, 36, 37
- Asian – 40, 41, 42, 43, 44
- Other ethnicity – 51, 52, 53, 61[49]

These are multiple assigned ethnic groups, in that a respondent can be assigned to multiple groups. For example, if a respondent reported being Māori and European ethnicity, they are assigned to both categories. The decision to code 62 to 'European' was to maintain comparability over time, and this decision was made in 2006 due to an assessment that this was the most reasonable way to treat such responses. The proportion of respondents with ethnic code 62 was 0.8% in 2014, hence whether this group was assigned to 'European' or 'Other ethnicity' currently makes little practical difference.

[48] There were 18 respondents included in the Māori booster sample dataset who did not select 'Māori' as one of their ethnicities. These respondents were sequenced to Q151_457 to check whether Māori was in-fact one of their ethnic groups and all confirmed their ethnicity included Māori hence remained as part of the final booster sample dataset. These respondents' ethnicity was back-coded by the ministry to include Māori.

[49] For detail on the 2-digit ethnicity codes see the Statistics NZ Level 2 classification http://www.stats .govt.nz/methods/classifications-and-standards/classification-related-stats-standards/ethnicity.aspx

Household composition

The question used to derive household composition (Q148_101) was modified for the 2014 NZCASS to align to the census living arrangements question. A respondent selected from a specified list the relationships they had to all people living in their household, including an 'Other – Specify' option where the text description of the relationship was collected.

The project team coded these responses into household composition categories. This process involved:

- determining appropriate household composition output categories from the statistical standard, with input from Statistics NZ
- conceptually mapping living arrangements to household composition, with review from Statistics NZ
- coding of 'Other' text responses
- deducing the number of parents for selected living arrangements
- manually coding the more complex living arrangements (approximately 10%) – that is, those who selected an 'Other' response
- coding in SAS the remaining respondents
- validation of output.

'Other – Specify' responses

A number of questions in the NZCASS questionnaire allowed the respondent to provide an 'Other – Specify' response. These were generally not back-coded, but in cases when the 'Other' group was reasonably large, these responses were investigated and qualitatively described in the Main Findings Report.

Geographic derivations

Three geographic data items were merged onto the NZCASS datasets for analysis:

1. the New Zealand Deprivation Index
2. urbanisation
3. regional classifications.

The New Zealand Index of Deprivation

The New Zealand Index of Deprivation 2013 (NZDep2013) was obtained from the University of Otago[50] and merged onto the NZCASS datasets by meshblock. There were 18 respondents from 3 meshblocks with a withheld NZDep due to technical reasons – described in the NZDep2013 report. The deciles were converted to quintiles through combining deciles 1 and 2, 3 and 4, etc.

Urbanisation

The Urban Area Classification 2013 was obtained from Statistics NZ and merged onto the NZCASS datasets by meshblock. The 3-digit urban area code was assigned to output categories as follows:

- main urban area – code range 001 to 100
- secondary urban area – code range 101 to 200
- minor urban area – code range 201 to 500
- rural – code range 501 upwards.

Regional classifications

The Region Classification 2013 was obtained from Statistics NZ and merged onto the NZCASS datasets by meshblock. The region codes were assigned to output categories as follows:

- Auckland = region code 2
- Wellington = region code 9
- Rest of North Island = region codes 1, 3, 4, 5, 6, 7, 8
- Canterbury = region code 13
- Rest of South Island = region codes 12, 14, 15, 16, 17, 18.

[50] http://www.otago.ac.nz/wellington/departments/publichealth/research/hirp/otago020194.html

Offence groupings

Offences often need to be grouped together rather than output as individual offence codes. For example, the 2 burglary offence codes (11 and 41) need to be grouped together to analyse burglary victimisation, or the number of abduction/kidnapping incidents is too small to output as a separate category so it is grouped together with assaults.

Offences are grouped together in different ways for different purposes. The 2014 NZCASS project team undertook the following process to determine how offences were to be grouped together:

- reviewed the way offences were grouped together for the 2009 NZCASS
- proposed a set of 5 offence groupings to be used for 2014 reporting, with consistent naming and labelling
- sought stakeholder feedback on whether proposed groupings meet their current needs, balanced with what is possible due to sample sizes.

Table 9.2 presents this standard set of 5 offence groupings used throughout 2014 NZCASS reporting (with the associated offence codes in brackets).

The grouping of offences was based on both the primary and secondary offence codes (if a secondary offence code was assigned). To improve alignment to New Zealand Police statistics, 1 exception was applied when burglary (11 or 41) was double coded with:

- 15 (unlawful interference/getting into motor vehicles)
- 16 (theft from motor vehicles)
- 17 (taking/conversion/unlawful interference with bicycle)
- 26 (damage to motor vehicles).

In these instances, only the burglary offence code was included.

This is a change since previous iterations of the NZCASS involving analysis of reporting to Police and victims' experiences and needs. For this analysis in previous years, only the primary offence code was included. These changes were applied retrospectively to the 2006 and 2009 NZCASS datasets for comparability.

Table 9.2: Offence groupings used in analysis

	Grouping 1 All offences – detailed	Grouping 2 All offences – broad	Grouping 3 Type of violence	Grouping 4 (subset of Grouping 5) Comparisons with Police data	Grouping 5 All offences – reporting to Police
Personal offences	Sexual offences (1, 2, 4)	Violent interpersonal offences[51] by relationship to offender [52]	Sexual offences (1, 2, 4)		Sexual offences (1, 2, 4)
	Threat – of force (21)		Threats and damage offences [53]		Threats (21, 29)
	Threat – to damage property (29)				
	Damage – personal property (28)				*Damage/theft – personal property (18, 28)
	Assault (6, 7, 8)		Physical offences (6, 7, 8, 9)	Assault (6, 7, 8)	Assault (6, 7, 8)
	Robbery (9)			Robbery/theft – from the person (9, 10)	Robbery/theft – from the person (9, 10)
	Theft – from the person (10)	Thefts and damage offences[54]			
	Theft – personal property (18)				*Damage/theft – personal property (18, 28)
Household offences	Theft – household property (12, 13, 17)				Damage/theft – household property (12, 13, 17, 19, 27)
	Damage – household property (19, 27)				
	Burglary (11, 41)	Burglary (11, 41)		Burglary (11, 41)	Burglary (11, 41)
	Theft – vehicle (14)	Vehicle offences (14, 15, 16, 26)		Theft – vehicle (14)	Theft – vehicle (14)
	Damage – vehicle (26)				Damage – vehicle (26)
	Thefts – from vehicle/vehicle interference (15, 16)			Thefts – from vehicle/vehicle interference (15, 16)	Thefts – from vehicle/vehicle interference (15, 16)

*Denotes the offence grouping with an asterisk is the same grouping but are not displayed side -by-side.

[51] 'Violent interpersonal offences' are defined as offence codes 1, 2, 4, 6, 7, 8, 9, 21, 29 and/or (27, 28 classified as 'directed' - see Figure 9:1). For selected analysis, this grouping was merged to be violent interpersonal offences and non-violent interpersonal offences.

[52] The hierarchy of relationship to offender is provided later in this chapter.

[53] 'Threats and damage offences' are defined as offence codes 21, 29 and/or (27, 28 classified as 'directed' – see Figure 9:1).

[54] 'Thefts and damage offences' are defined as offence codes 10, 18, 12, 13, 17, 19 and/or (27, 28 classified as 'non-directed' – see **Error! Reference source not found.**).

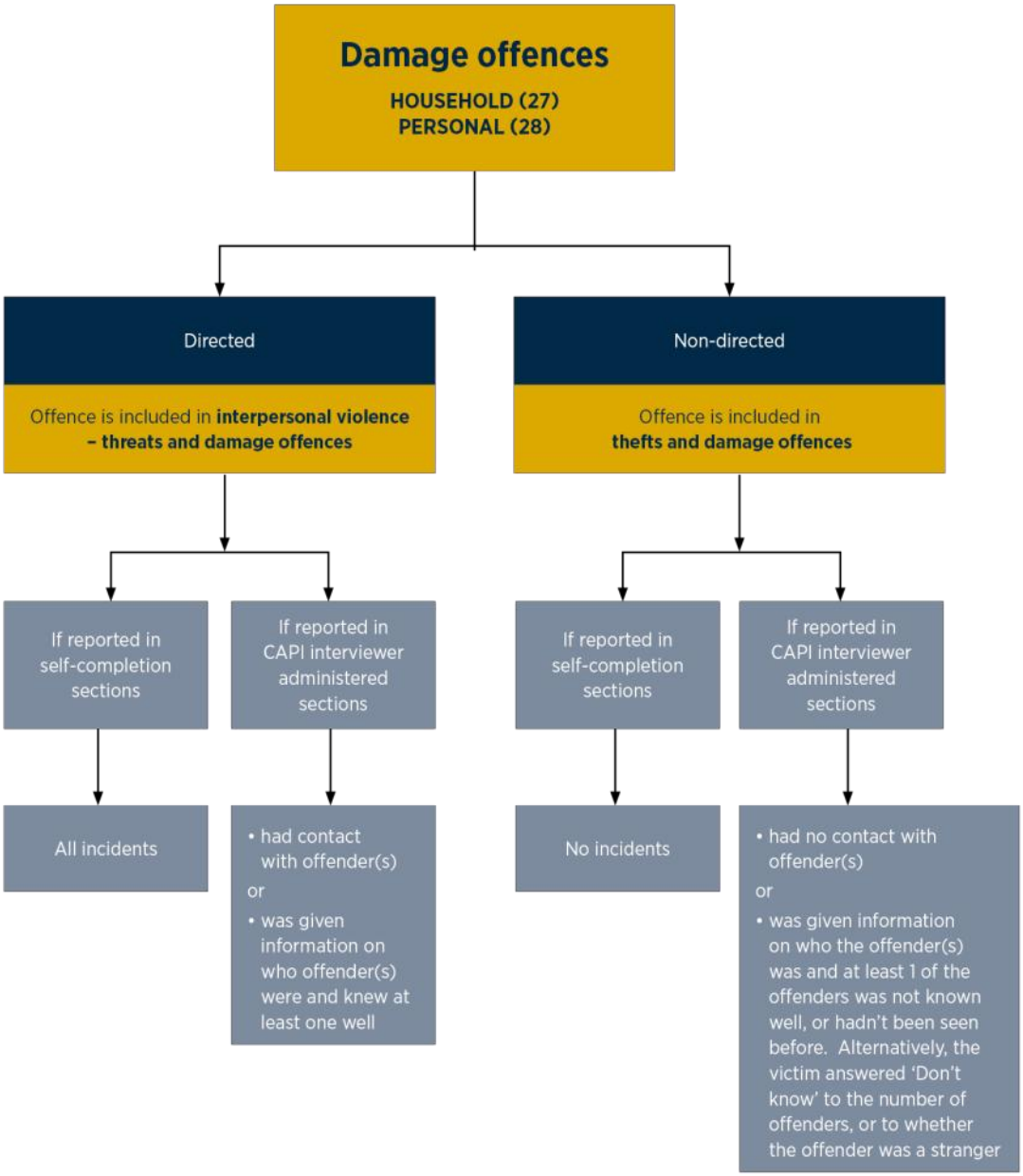
Separating damage offences

As presented in Table 9.2, household and personal damage offences (offence codes 27 and 28 respectively) were classified into either:

- 'threats and damage offences' (as part of interpersonal violence); or
- 'thefts and damage offences'.

The criteria used to separate these offences are presented in Figure 9.1.

Figure 9.1: Separation of damage offences



Interpersonal violence groupings

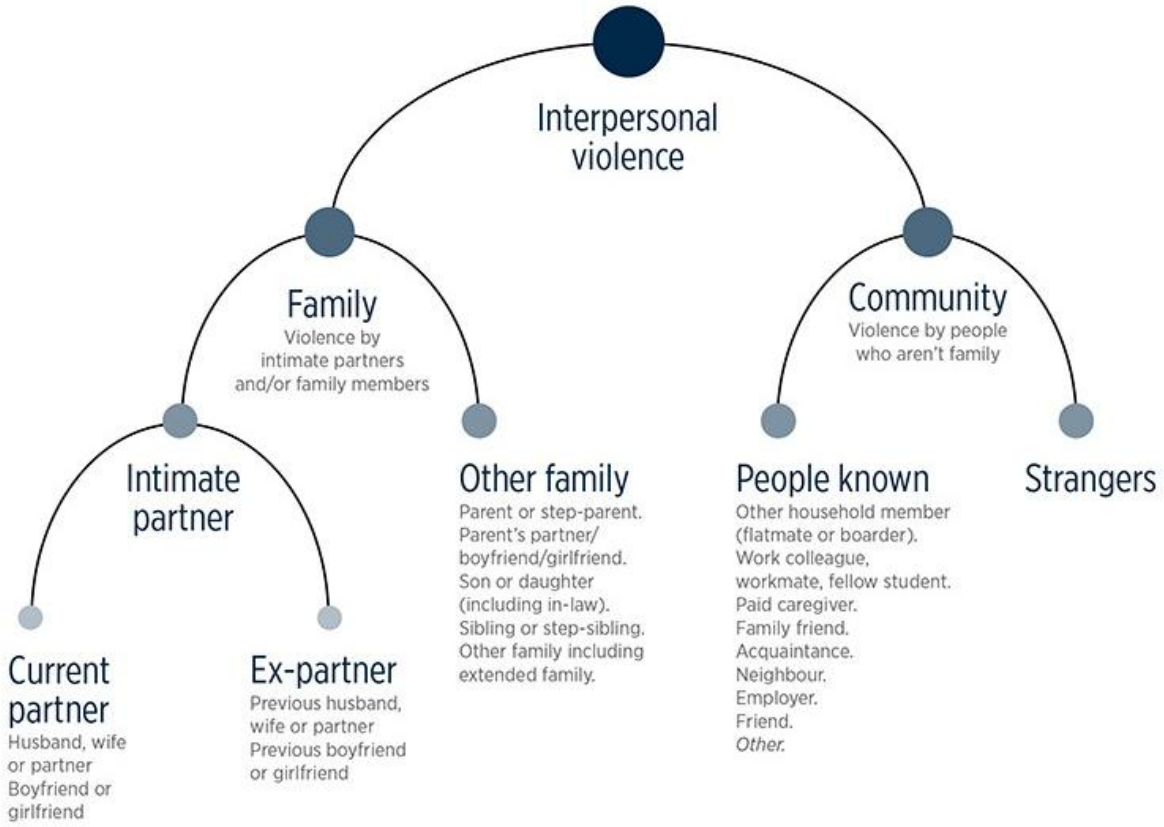
The framework used to report on violent interpersonal offences was reviewed in 2014 in order to better meet stakeholder needs. The groups in the NZCASS interpersonal violence reporting framework are based on:

- the victim's relationship to the offender
- the type of violence experienced.

Victim's relationship to the offender

Where a victim had contact with the offender or came to know who committed the offence, they are asked: 'What were their relationships to you at the time it happened?' This information is used to group relationship types as shown in Figure 9.2.

Figure 9.2: Interpersonal violence relationship to offender framework



In reporting, the group used for analysis largely depends on the sample size. For example, if the sample is too small to look at estimates for 'Intimate partner' and 'Other family', analysis will be done at the next level in the hierarchy – 'Family' (violence by intimate partners or other family members).

Some of the relationships included in the groups above will differ from some other countries. For example, boyfriends and girlfriends have been included in the 'Intimate partner' groups for the NZCASS but are often not part of definitions of intimate partners in similar Australian statistics.

Type of violence experienced

The groupings used in the NZCASS are as follows: physical violence, sexual violence, threats and damage. Table 9.3 shows the NZCASS offences that are in or out of scope for each of these groups.

Table 9.3: Types of interpersonal violence

Type of violence	In scope	Out of scope[55]
Physical violence and abuse	• assault • kidnapping and abduction • robbery	
Sexual violence and abuse	• sexual violation • indecent assault	• indecent exposure (collected in the NZCASS but not counted in statistics)
Psychological violence	• threats • damage to property – personal • damage to property – household (when the victim had contact with the offender, or if the victim was given information about who the offender was and knew them well)	• Deprivation/Neglect • Coercive & Controlling behaviours[56]

[55] Reminder: Children under 15 years old and those living in institutions (such as aged care homes) are out of scope for the NZCASS.

[56] Some coercive & controlling behaviours are collected as part of the NZCASS however, these are collected differently from offences.

Household composition groupings

Table 9.4 provides descriptions of each household grouping.

Table 9.4: Household composition groupings descriptions

Composition grouping	Description
One-person household	Lives alone.
One parent with child(ren)	One person living with their son(s) and/or daughter(s) (natural, step, adopted or foster).
One parent with child(ren) and other person(s)	This household could include another person that is unrelated, such as a flatmate or boarder or could be related but not part of the immediate family unit, such as parent's sibling/children's aunt.
Couple only	Two persons who are either opposite-sex or same-sex spouses/civil union partners/partners.
Couple with no children and other person(s)	This household could include another person, such as a flatmate, boarder or a family member, such as a parent of 1 couple member.
Couple with children	Two persons who are either opposite-sex or same-sex spouses/civil union partners/partners, living with their son(s) and/or daughter(s) (natural, step, adopted or foster).
Couple with children and other person(s)	This household could include another person that is unrelated, such as a flatmate or boarder or could be related but not part of the immediate family unit, such as parent's sibling/children's aunt.
Multiple family household	This is when multiple families are living in the same household – for example, 2 married couples flatting together or a married couple plus 1 partner's mother and father (which is considered a second family unit).
Other multi-person household	This comprises households of related and/or unrelated people, where there are no couples or parents with a child. It consists, for example, of flatting arrangements, 2 siblings living together or 1 person with a boarder.

Note: The terminology of 'children' can relate to young children or adult children, but are defined to be children if they do not have their own partners or children of their own living in the household.

10. Weighting

Introduction

This chapter describes the methods used to produce weights and replicate weights for the 2014 NZCASS.

Weights are usually applied to sample survey data during its analysis to adjust for factors such as differential selection probabilities, non-response patterns and sample skews relative to population figures. The 2014 NZCASS is no exception.

The sample design for the 2014 NZCASS incorporated 4 levels: meshblocks, households, people, and victimisation incidents. Weights have been calculated to enable analysis of the NZCASS data at 3 of these levels: households, people, and incidents. These weights incorporate adjustments for each of the factors listed above.

Changes to weighting

The weighting process for the 2014 NZCASS is very similar in general to that used for the 2009 NZCASS. Some changes were required to accommodate changes in the sample design, most importantly the selection of booster sample addresses from the electoral roll, but also including the removal of an upper limit on the number of main sample addresses approached per meshblock, and the increased number of booster sample selections in the last 2 months of 2014 fieldwork. Population benchmark figures were also updated.

Ethnicity in weighting

There were 18 respondents included in the Māori booster sample dataset who did not select 'Māori' as one of their ethnicities in Q151. These respondents were sequenced to Q151_457 to check whether Māori was in-fact one of their ethnic groups. These 18 respondents confirmed their ethnicity included Māori hence remained as part of the final booster sample dataset. The initial 2-digit ethnicity codes for these respondents did not include Māori, but they were later back-coded by the ministry to include Māori. This means all NZCASS analysis has correctly classified ethnicity.

Because ethnicity is used in the weighting process and weights were derived prior to the ethnicity codes being corrected, it means these 18 respondents were not classified as Māori (but were classified as the other ethnic groups they selected) in the weighting process.

The scale of the impact on changes to weights was assessed on all headline NZCASS statistics, and the most detailed Māori table. This process showed that the changes were minor and none of the differences were statistically significant.

Household weights

Initial household weights were calculated as the reciprocal of each household's estimated probability of inclusion in the sample, across both the Māori booster sample and the main sample. (Person weights were calculated using a similar process.) The inclusion probability for household i was calculated as follows:

$$\begin{aligned} p_i &= P(\text{household } i \text{ included}) \\ &= P(\text{household } i \text{ selected and eligible}) \\ &= \sum_{s-1}^{2} p_{i,s}\, r_{i,s} \end{aligned}$$

where $p_{i,s}$ is the probability that household i was selected for sample s and $r_{i,s}$ is the probability that household i was eligible for sample s. The main sample and Māori booster sample are indexed by s = 1 and 2 respectively.)

One thousand meshblocks were selected systematically with unequal probabilities that were proportional to the number of occupied private dwellings they contained according to the 2013 Census.[57] So the probability of selecting a meshblock m that contained d_m occupied private dwellings according to the 2013 Census is $1{,}000d_m/D$, where D was the total number of private occupied dwellings in the sampling frame according to the 2013 Census.

Residential addresses in each selected meshblock m were extracted from NZ Post's PAF. Certain addresses from the electoral roll were available[58] and were added to this list. A fraction of the listed addresses were selected systematically (ie every $x_{m^{th}}$ address was chosen as part of the main sample). Here x_m was calculated as the number of occupied private dwellings the meshblock contained at the 2013 Census (denoted d_m), divided by 6.8571,[59] then randomly rounded to an adjacent integer. The probability that 1 of the listed households within meshblock m was selected for the main sample was therefore $1/x_m$. The interviewer also enumerated all occupied private dwellings that were absent from the original list of addresses in the meshblock when they first visited the meshblock. Any such dwellings were added to the list, and the same fraction $1/x_m$ of these additional enumerated dwellings was selected.

As a result, the probability $p_{i,s}$ that each occupied private dwelling was selected for the main sample was:

$$\begin{aligned} p_{i,1} &= P(\text{household } i \text{ selected for main sample}) \\ &= P(\text{meshblock } m \text{ selected}) P\left[\begin{array}{l}\text{household } i \text{ selected for main sample} \\ \text{within meshblock } m\end{array}\right] \\ &= \frac{1{,}000 d_m}{D}\frac{1}{x_m} \end{aligned}$$

[57] Meshblocks with fewer than 9 occupied private dwellings were excluded from the sampling frame, as were meshblocks in inlets, waterways, and on islands other than Waiheke Island and the North and South Islands.

[58] A list of registered electors of Māori descent, including their addresses, was obtained to enable more effective booster sampling.

[59] This figure (6.8571) is the average number of dwellings to be selected for the main sample per meshblock (excluding enumerated dwellings), derived by dividing the target main sample size (4800) by the number of meshblocks selected (1000) and the expected response rate (70%).

whether they were on the combined PAF/electoral roll list or were enumerated later. Most occupied private dwellings thus had approximately the same probability of being approached for the main sample. This is similar to the outcome of the sample designs used in 2006 and 2009 surveys.

Dwellings were also selected for the Māori booster sample within the meshblocks selected for the main sample. They were systematically sampled from addresses on the electoral roll containing electors of Māori descent, excluding addresses already selected for the main sample. The number of booster sample addresses to approach in each meshblock was calculated so as to be proportional to the total number of such addresses in the meshblock, then randomly rounded to an adjacent integer. The number of booster addresses to approach was initially calculated assuming that 60% of addresses would yield an interview, but as fieldwork proceeded it became apparent that this assumption was unrealistic. For meshblocks with interviewing scheduled in May and June, the numbers to be approached were revised to increase the booster sample by 700 addresses in total, reflecting an expected yield of around 48%. Since meshblocks were assigned to months essentially at random, booster sample addresses can be treated as having a uniform selection probability across all months.

The probability $p_{i,2}$ that an address was selected for the Māori booster sample would be zero if it was not listed on the electoral roll as containing any electors of Māori descent. Otherwise it was:

$$\begin{aligned} p_{i,2} &= P(\text{household } i \text{ selected for booster sample}) \\ &= P(\text{meshblock } m \text{ selected}) P\left(\begin{array}{l}\text{household } i \text{ selected for booster} \\ \text{sample within meshblock } m\end{array}\right) \\ &= \frac{1{,}000 d_m}{D} \frac{0.6 b_{1,m} + 0.4 b_{2,m}}{A_m} \end{aligned}$$

where A_m is the number of addresses listed as containing electors of Māori descent within meshblock m, not counting addresses that had been selected for the main sample; $b_{i.m}$ is the number of addresses that would be selected for the booster sample in meshblock m if it was scheduled for interviewing in February to April; and $b_{2,m}$ is the number of addresses that would be selected for the booster sample in meshblock m if it was scheduled for interviewing in May or June.

This formula for the household selection probabilities differs from the corresponding formulae for the 2006 and 2009 surveys. The differences reflect the changes in the surveys' sample designs, especially for the booster sample. Analysing the data using the resulting survey weights will ensure that these differences do not materially affect comparisons between the 2 surveys' results.

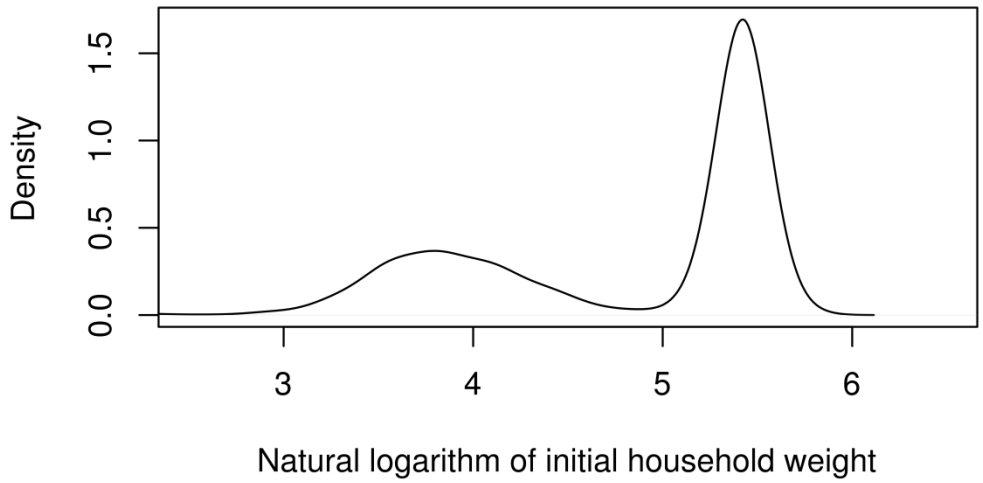
Eligibility for the Māori booster sample was not collected in the 2006 NZCASS for non-Māori respondents in the main sample. In that survey, the probability of eligibility for the Māori booster sample $r_{i,2}$ was estimated for these respondents as $r_{i,2} = \frac{uv}{1-(1-u)v}$ where u is the proportion of people in the household aged 15 or over who would be eligible for the Māori booster sample, averaged over all households known to be of mixed eligibility, and v is the proportion of households that were eligible. For comparing later surveys with the 2006 NZCASS, an alternative household selection probability was computed using the same approach used for the 2006 NZCASS, using the 2009 NZCASS data to estimate v. Specifically, v = (1297 + 471)/(1297 + 11554 + 471) = 0.13271.[60] In the 2014 NZCASS, u = 0.11908, and thus for non-Māori living with other adults, $r_{i,2}$ = 0.01789.[61]

[60] In the 2009 NZCASS, 1297 booster interviews were conducted, another 471 households were approached for the booster sample and were eligible, while 11,554 households were approached for the booster sample but were ineligible.

[61] The value of ri,2 is a constant for each survey; this means that it applies unchanged to all non-Māori living with other adults in that survey.

The initial household weight was calculated as the reciprocal of the household inclusion probabilities described above. The resulting weights had an average value of 159.8, a coefficient of variation of 0.547, and ranged from 7.1 up to 312.4. They were noticeably more variable than in the 2009 NZCASS (as would be expected from the changes to the sample design). The estimated probability density of the logarithm of the initial household weights is shown in Figure 10.1.[62] The distribution in Figure 10.1 is bimodal because the Māori booster sample gives Māori a higher chance of being selected.

Figure 10.1: Density of the logarithm of the initial household weight



A non-response adjustment was made to these initial household weights, to allow for differential unit non-response. This adjustment was expected to be especially useful for variables for which population totals are not available, such as which sample the household was part of, because these cannot be adjusted for in the later post-stratification step. It does assume that the missing data is missing at random (conditional on the adjustment variables). If this is not a realistic assumption, survey results could be biased, as could comparisons between surveys.[63]

To adjust for non-response, the response outcome data was summarised using a single overall logistic regression model. This was fitted to the dataset containing the outcome counts by samples for all 1000 meshblocks, with the number of trials in each meshblock taken as the estimated number of eligible dwellings approached (rounded to the nearest whole number), and interviews being counted as successful trials. First a model was fitted using the following predictor variables: sample (Māori booster or main sample), the 2013 crime rate in that meshblock (on a truncated log scale), 2013 broad region, 2006 level of urbanisation[64], and deprivation index (NZDep13). Then the non-response model was selected by sequentially removing predictor variables that were not statistically significant, until only significant variables remained at a 90% confidence level.

---

[62] The density shown in Figure 10:1 was estimated using a Gaussian kernel, with the bandwidth given by the rule of thumb in equation 3.3.1 on page 48 of Silverman (1986). That is, the bandwidth was the minimum of the interquartile range divided by 1.34 and the standard deviation, multiplied by 0.9 divided by the sample size to the power of one-fifth.

[63] The term 'bias' is used here in a technical sense, meaning the extent to which the average of the results would not agree with the true population figures (if these were known), supposing that the survey was conducted many times in the same circumstances but different samples were selected according to the same design.

[64] 2006 level of urbanisation was used as it was the most recent available at the time of the weighting process.

The final predictors were the sample, level of urbanisation, and broad region – their parameter estimates are shown in Table 10.1. Initial household weights were divided by the predicted probabilities of response based on this model, which ranged from 0.749 to 0.914. The resulting weights had an average of 197.5 and a coefficient of variation of 0.559, and ranged from 8.8 to 397.5.

Table 10.1: Parameter estimates for non-response model

Variable	Parameter estimate	Standard error
Intercept	1.0906	0.0514
Sample – booster	0.3368	0.0796
Main urban areas excluding Auckland, Wellington, Christchurch and Dunedin	0.3713	0.0813
Satellite and independent urban areas	0.2597	0.0904
Rural areas	0.4747	0.1028
Lower North Island	0.4588	0.0842
South Island	0.2820	0.0811

The household weights resulting from the non-response adjustment were then post-stratified[65] by level of urbanisation[66], based on the estimated number of households in each category as at 30 June 2014 (see Appendix E). This aligned the total of the household weights in each urbanisation category with the estimated number of households shown in Table 10.2 below.

Table 10.2: Population targets for household weight calculation

Urbanisation	Number of households
Auckland	469,048
Wellington/ Christchurch/Dunedin	360,614
Other main urban areas	356,219
Other urban areas	259,208
Rural areas	250,710

The household weights were post-stratified by urbanisation for consistency with earlier NZCASS iterations, and because urbanisation is generally associated with both crime rate and non-response. There were few other good benchmark candidates for households.

The final household weights after post-stratification ranged from 10.5 to 509.7, with an average of and a coefficient of variation of 0.561. These weights can be used for analyses of household characteristics, and in particular to calculate incidence and prevalence figures for household offences.

[65] Post-stratification is a widely used technique for adjusting survey weights so that the results agree with known population benchmarks.

[66] The urbanisation categories used in weighting differ from the urbanisation categories used to output the 2014 NZCASS data. The output categories were updated as part of the 2014 NZCASS to better align to standard classifications, whereas the categories used in weighting were retained for consistency with the 2006 and 2009 NZCASS weighting process. Total weights for urbanisation categories are not reported for NZCASS. Hence the only impact is a minor increase in variance for victimisation counts (not rates) by urbanisation (which is not part of standard NZCASS output).

Person weights

Person weights were calculated using a similar process to that described above for the household weights, with initial inverse probability weights being adjusted for unit non-response and then aligned with population benchmarks. The only differences were that the selection probabilities incorporated an extra factor to account for the selection of 1 person from those in the household who were eligible to be interviewed, and that more than 1 benchmark variable was used, requiring the use of raking[67] instead of post-stratification.

The components of the household inclusion probabilities for each sample were divided by the number of people living in the household who were eligible to be interviewed[68] (to adjust for only 1 person from each household being interviewed), according to the following formula:

$$o_i = P\ (\text{person } i \text{ was included})$$
$$= \sum_{s-1}^{2} \frac{p_{i,s} r_{i,s}}{e_{i,s}}$$

where $e_{i,s}$is the number of people aged 15 or over living with respondent i who were eligible for samples, except that $e_{i,2}$ is taken as zero for non-Māori respondents. Initial person weights were taken as the reciprocal of each person's inclusion probability (ie as $1/o_i$).

Adjustment for unit non-response used the same non-response model as for households (ie the initial person weights were divided by the same predicted probabilities of response as for the household weights). The person weights were then raked by combinations of age,[69] sex and ethnicity. These combinations are consistent with those used in the 2006 and 2009 surveys.[70] Weighted sample profiles (before raking) show substantial skews relative to the population benchmarks for several of these groups, especially those relating to Māori, and none of the groups have small sample sizes. (The smallest group was Māori males aged 60 or more, which contained 198 respondents.) The population targets used are shown in Table 10.3, in the order that they were raked.

The initial inverse probability person weights ranged from 7.1 to 1562.0, with an average of 339.4 and a coefficient of variation of 0.757. After the non-response adjustment, the person weights had an average of 421.1 and a coefficient of variation of 0.776. The final person weights after raking ranged from 12.4 to 2773.0, with an average of 510.3 and a coefficient of variation of 0.770.

Person weights can be used in the calculation of incidence and prevalence figures for personal offences, and for the analysis of self-completion lifetime prevalence data and of most data from the CAPI section. No further adjustments have been made to account for non-response to the entire self-completion component, on the grounds that this is consistent with previous iterations of the NZCASS and because non-response here remains small.

---

[67] Raking, also known as rim weighting, enables the simultaneous control of marginal distributions for several benchmark variables (in contrast to post-stratification, which controls only on a single categorical variable). It was implemented here using Lumley's (2004, 2013) rake function, with the default convergence criterion, and that function's source code provides the most precise description of the method used.

[68] If the number of eligible household members was greater than 6, a value of 6 was used instead. This affected 27 respondents.

[69] Six respondents refused to give their age. The resulting missing values were imputed using random hot deck imputation.

[70] As is the case with urbanisation discussed earlier, the age groups used in weighting differ from the age groups used to output the 2014 NZCASS data. The output age groups were updated as part of the 2014 NZCASS to better align to standard classifications, whereas the age groups used in weighting were retained for consistency with the 2006 and 2009 NZCASS weighting process. See the urbanisation footnote for how this impact statistics.

Table 10.3: Population targets for person weight calculation

Weighting control group	Estimated population at 30 June 2014
Males 15–24	327,760
Males 25–39	406,820
Males 40–59	587,550
Males 60–69	225,170
Males 70+	195,210
Females 15–24	311,470
Females 25–39	435,070
Females 40–59	634,320
Females 60–69	236,320
Females 70+	238,920
Māori Males 15–39	122,920
Māori Males 40–59	70,310
Māori Males 60+	28,120
Māori Females 15–39	132,490
Māori Females 40–59	80,370
Māori Females 60+	33,030
Non-Māori	3,131,370
Pacific	226,535
Non-Pacific	3,372,075
Asian	441,543
Non-Asian	3,157,067

Incident weights

Incident weights were derived from person weights by dividing them by the selection probability for that incident (given that the current respondent had been selected). If the incident selection probability was less than 0.1, however, a value of 0.1 was used instead.[71] This reduced the variability of these probabilities, and thus reduced the variability in the weights that would result from using the actual incident probabilities from heavily victimised respondents, although this is at the cost of introducing some potential for bias. Since a value of 0.1 was used in the 2006 and 2009 surveys, it was used again here for consistency. The same cut-off value was used for incidents from both the CAPI section and the CASI section.

The most recent incident was selected within each part of the CASI section, but a more complex selection process was used for incidents recorded in CAPI section. These were divided into high, medium or low priority incidents,[72] and an incident was selected without replacement for each of the 3 victim forms in turn. High priority incidents were given 3 times more chance of being selected for each victim form than a low priority incident, and medium priority incidents were given twice as much chance. The overall probability that a given incident from the CAPI section was selected for a victim form thus depended on whether it was a high, medium or low priority incident, as well as how many low, medium and high priority incidents were experienced by that participant.

The joint probability of selecting various numbers of high, medium or low priority incidents was derived by enumerating the relevant parts of the probability space through a branching process. The probability of selecting an incident was then calculated adding up the probabilities for each of the appropriate nodes. For instance, if a respondent reported 1 low priority incident, 1 medium priority incident, and 2 high priority incidents, the probability that a low priority incident was selected for the first victim form would be 1/9, and the probabilities of a medium and high priority incident being selected would be 2/9 and 6/9 respectively. (Note: This is the probability of any high priority incident being selected, not a specific incident.) If we suppose the low priority incident was the one selected, then 1 medium priority and 2 high priority incidents would remain, and the conditional probability of selecting a medium priority incident for the second victim form would be 2/8. The joint probability of selecting a low priority incident for the first victim form and a medium priority incident for the second victim form is the product (1/9) × (2/8) = 2/72. Similarly, the conditional probability of selecting a high priority incident for the second victim form would be 6/8, and the joint probability of selecting a low priority incident for the first victim form and a high priority incident for the second victim form would be (1/9) × (6/8) = 6/72.

Similar calculations give the full joint distribution for the priority of the incidents selected for the first 2 victim forms as:

$$\begin{bmatrix} 0 & 2/63 & 6/54 \\ 2/72 & 0 & 12/54 \\ 6/72 & 12/63 & 18/54 \end{bmatrix} = \begin{bmatrix} 0 & 0.0317 & 0.1111 \\ 0.0278 & 0 & 0.2222 \\ 0.0833 & 0.1905 & 0.3333 \end{bmatrix}$$

where the first victim form indexes the columns, and the second the rows.

Now suppose that a low priority incident was selected for the first victim form, and a high priority incident was selected for the second victim form. Then 1 medium priority and 1 high priority incident would remain, and the conditional probability that a medium priority incident would be selected for the third victim form would be 2/5. The unconditional joint probability of selecting a low priority

[71] This truncation of the distribution of probabilities affected 133 incidents. Although these incidents made up only 3% of sampled incidents, they would have accounted for 62% of the total of all incident weights if this truncation had not been applied. After truncation, they accounted for only 12% of the incident weight total.

[72] High priority incidents were those recorded at questions 28, 36, 37, 38 and 40. Medium priority incidents were those recorded at questions 31, 35, 35.416, 39 and 41. Low priority incidents were those recorded at questions 29, 30, 32, 34 and 43.

incident for the first victim form, a high priority incident for the second, and a medium priority incident for the third victim form would be (6/72) × (2/5) = 0.03333.

Similar calculations give the full joint distribution for the priority of the incidents selected for all 3 victim forms as:

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0.0556 \\ 0 & 0.0467 & 0.1111 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0.0444 \\ 0 & 0 & 0 \\ 0.0333 & 0 & 0.2222 \end{bmatrix}, \begin{bmatrix} 0 & 0.0317 & 0.0667 \\ 0.0278 & 0 & 0.1667 \\ 0.05 & 0.1429 & 0 \end{bmatrix}$$

where the third victim form indexes across matrices (and the first victim form indexes the columns, and the second victim form indexes the rows, as before). Summing the entries where a low priority incident was selected (ie the first matrix, and the first row and column of the other 2 matrices) gives the total probability that a low priority incident was selected for any of the 3 victim forms as 0.468254.

The CASI section comprised 3 parts: incidents committed by partners; incidents committed by people the respondent knew well; and sexual incidents. Each part included 4 screener questions establishing how many incidents had occurred, followed by a series of questions in a victim form (which corresponded to questions in the victim forms in the CAPI section). If more than 1 incident was reported in a part, the corresponding victim form questions were completed for the most recent incident. Although this was not a random selection, it was assumed for weighting purposes that the incident was selected at random, with equal probability given to all the incidents reported in that part.[73] The resulting selection probability was the reciprocal of the number of incidents reported at all screener questions in that part.

Isolated missing values for the number of incidents (eg from a 'Don't wish to answer' response to a particular screener question) were imputed with the value 1, as was done in 2006 and 2009.

While it might seem more consistent to derive a second set of incident weights from the household weights rather than the person weights, and use these for analysing household incidents, this would introduce some bias in comparisons against previous results because incident weights were derived from person weights for all incidents in the 2006 and 2009 surveys. This previous approach produces results in terms of person-incidents; that is, the numbers of incidents experienced by people living in the affected households. This better matches the common interpretation of the results in terms of the proportions of victims affected.[74] To maintain comparability with the previous approach, person weights have again been used for all incidents in the 2014 NZCASS.[75] If the weighted results are interpreted in terms of incidents, this introduces some bias.[76] It is more correct to interpret the weighted results in terms of person-incidents, as explained above.

The incident weights ranged from 15.1 to 19540.0, with an average of 874.5 and a coefficient of variation of 1.51. The density plots for the weights in Figure 10.2 show that incident weights are more variable than person weights, which in turn are more variable than household weights. This reflects the large variation in selection probabilities at each of the last 2 stages of the 2014 NZCASS sample design.

[73] The assumption of random equiprobable selection of CASI incidents within each selection follows from a model that assumes the interview date is not related to patterns of victimisation, and that respondents recall and report incidents perfectly. The latter assumption is already made implicitly elsewhere in the weighting and analysis.

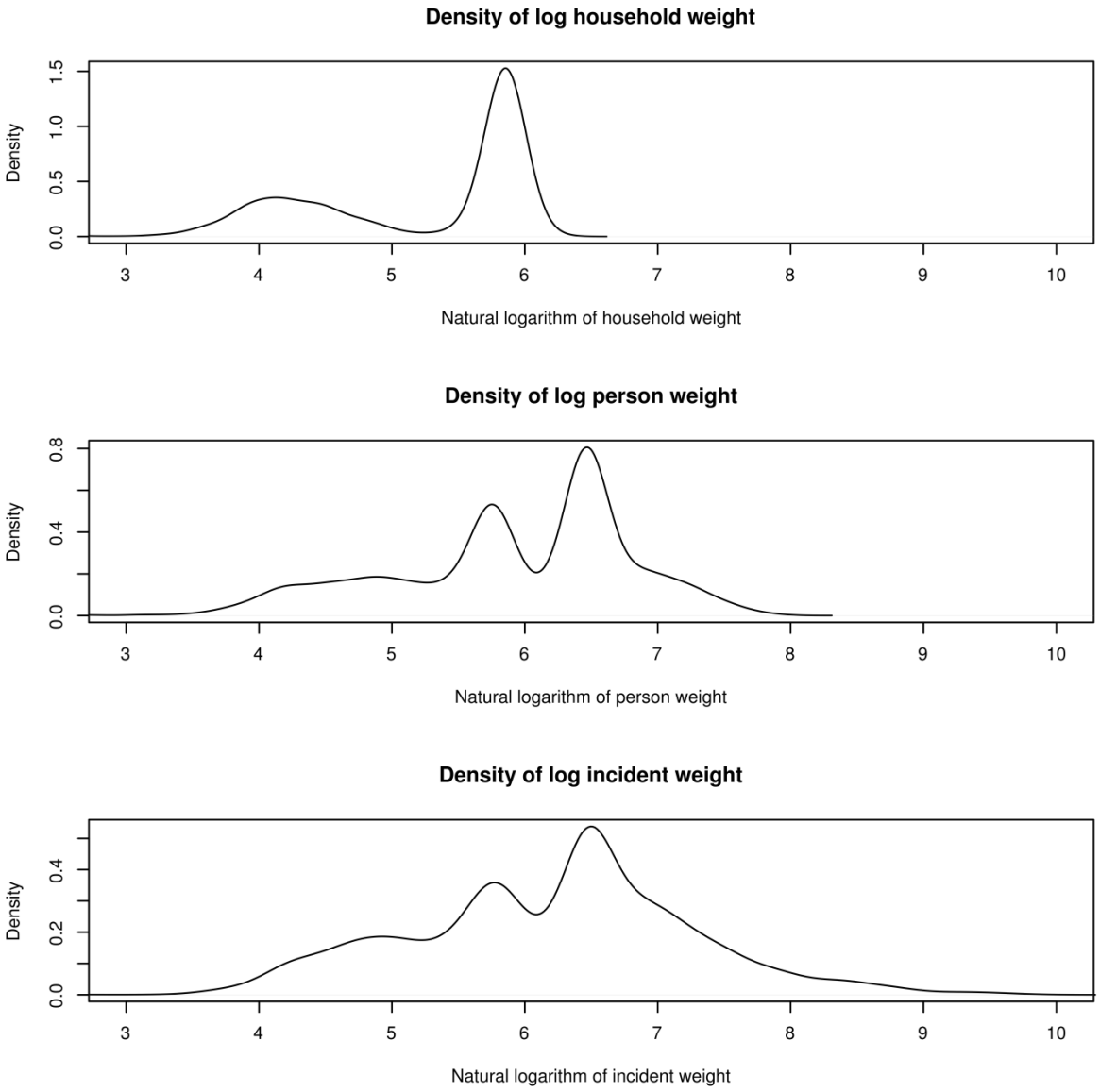
[74] The distinction between victims and incidents was explicitly discussed in footnote 35 on page 61 of the 2006 Key Findings Report, which explains that this distinction is generally not maintained within the report.

[75] Another option would be to derive household incident weights from the household weights and use these to produce the main estimates, but use incident weights derived solely from person weights for comparisons with previous surveys.

[76] This interpretational bias would be particularly severe if estimates of total numbers of incidents were produced from these weights, although in previous surveys all results have been expressed as percentages.

Incident weights were merged into the incident dataset for analysis purposes. This dataset contained a record for each victim form completed, and so could contain up to 6 records for each respondent—3 from the CAPI section and 1 from each part of the CASI section.

Figure 10.2: Densities of logarithms of the household, person and incident weights



Replicate weights

While sample surveys like the NZCASS provide a practical and cost-effective means of collecting information on victimisation, the survey results are inherently subject to random sampling variation. The size of this variation must be estimated and considered to sensibly interpret the results.

Variance estimation for the NZCASS is complicated by the large amount of missing data (covered in Chapter 11: Imputation) and the survey's complex sample design. Replicate weights were calculated using the delete-a-group jackknife method (Kott 1998) to accommodate the sample design and weighting for the 2014 NZCASS.[77]

The delete-a-group jackknife, like other resampling methods, uses the variation between the results for many sample 'replicates' to estimate sampling variances (excluding imputation effects).

Replicates were created by first randomly dividing the PSUs into equal groups, then omitting 1 group from the sample to form each replicate. Each replicate can equivalently be thought of as assigning the 'omitted' group zero weight (and increasing the weights for other respondents to compensate) instead of actually removing them from the dataset. For the 2014 NZCASS, 100 replicates were used. That is, the 1000 PSUs (meshblocks) were randomly divided into 100 groups of 10 meshblocks, each of which formed the omitted group for 1 replicate.

The weighting process was rerun[78] for each of the 100 jackknife replicates, producing 100 sets of replicate weights, to account for the effect of the weighting framework.

See Chapter 12 for how replicates weights are used to estimate variance.

Quality assurance

As part of the 2014 NZCASS, a line-by-line review of the weighting code for all three years (2006, 2009 and 2014) was undertaken by the ministry's independent quality assurance (IQA) provider from the statistics department of Auckland University.

This process was to ensure that the code was undertaking weighting as prescribed and was fit for purpose before the weighting was implemented and analysis was undertaken.

[77] In contrast, balanced repeated replication was used for the 2006 NZCASS. This technique can handle stratification and non-negligible sampling fractions well, but relies on imperfect workarounds for unrealistic assumptions such as each stratum containing exactly 2 sampled PSUs. Because the 2009 and 2014 samples were unstratified and their first stage sampling fractions were negligible, the simpler delete-a-group jackknife was used for the 2009 and 2014 surveys instead.

[78] However, the same non-response adjustment model was used across all replicates (ie model selection was not rerun for each replicate). This may lead to sampling errors being slightly underestimated.

11. Imputation

Introduction

Each respondent could only complete up to a maximum of 6 victim forms, due to interview time constraints. Since some respondents are heavily victimised, some incidents do not have a victim form completed. Victim forms were completed for only 17% of the incidents reported in the 2014 NZCASS. This is lower than in previous years, primarily due to more respondents reporting having experienced large numbers of incidents in the CASI section.[79] The effect of these large values is limited by truncation, as described in the "Heavy victimisation cut-off" section later in this chapter.

Table 11.1: Percentage of victim forms and respondents where imputation required, by year

	2006	2009	2014
Percentage of incidents reported where victim forms were completed	33%	37%	17%
Percentage of respondents reporting having experienced too many incidents for all to have victim forms	19%	15%	10%

For comparison, 10% of respondents reported too many incidents in the 2014 NZCASS for a victim form to be completed for all of them, i.e. they reported having experienced more than three incidents in the CAPI questionnaire, or more than one incident in any of the three CASI sections. This percentage has declined from 19% in 2006. Of those reporting any CAPI incidents in NZCASS 2014, 16% reported experiencing more than three CAPI incidents in total. Of those reporting any CASI incidents, 59% reported more than one incident in at least one of the CASI sections. The incident weights described in Chapter 10 adjust for this to provide an effective method of analysis when each victim form can be viewed in isolation. Most of the data collected in victim forms will be analysed using these weights.

Victimisation rates are also a critical output from the NZCASS. These are usually expressed as:

- prevalence rates (the proportion of people or households who experienced offences of a certain type)
- incidence rates (the average number of such offences experienced by a person or household).

[79] There were 61 values greater than 30 reported by respondents at the CASI screener questions gathering the number of incidents reported. A single respondent reported the maximum value (999 and 99 respectively) at each of the 8 screeners in the second and third CASI sections; these were not considered reliable and were instead treated as missing.

Direct calculation of incidence and prevalence rates required information about all the incidents experienced[80] by each survey participant, including 3 items collected on each victim form:

1. whether the incident occurred during the year 2013
2. whether the incident was determined during offence coding to be an offence within the scope of the survey (termed being relevant)
3. which detailed offence codes the incident attracts.

Another item that can be derived from victim form data (under certain assumptions) is how likely the incident was to have been reported a second time at another screener question.

Information on these items was required for all incidents, including the 83% of incidents without victim forms,[81] to enable the direct calculation of incidence and prevalence figures.[82] However, this data is missing for the incidents without victim forms, necessitating some form of imputation for this missing data.[83]

The process of imputation also affects the sampling error of the results, although for some imputation methods it is hard to figure out how much. Multiple imputation (Rubin 1987) has been used in the NZCASS to quantify this effect. In 2014, 100 imputations were used for analysis of all 3 years (2006, 2009 and 2014).

[80] This is most critical for prevalence rates, since the number of offences could be calculated directly from the incident weights, and thus so could incidence rates. These figures would be less reliable than those based on imputation, however. In contrast, prevalence rates cannot be calculated directly using weights. Prevalence is defined as the proportion of people (or households) who experienced the offence in question, but the available data does not establish which offences were experienced for people with missing victim forms; that is, who reported experiencing more than 3 incidents in the CAPI section (or more than 1 incident in any part of the CASI section). Imputation provides a natural solution to this problem, and since imputed values have been produced to enable calculation of prevalence rates, it makes sense to use the same values to calculate incidence rates.

[81] The proportion of incidents without victim forms depends on the questionnaire design, and in particular the choice to gather victim form data for just 1 incident from those reported in each part of the CASI section and at most 3 incidents from the CAPI section. One option for reducing the proportion of incidents without victim forms was a mini -victim form developed during planning for the 2006 NZCASS, to gather only the information required for calculating incidence and prevalence, for 1 more incident from the CAPI section. It was ruled out then due to constraints on interview duration, omitted from the 2009 NZCASS for consistency and not considered in 2014 as it was out of scope for the project.

[82] Alternative methods that do not require data for all incidents are theoretically possible, such as pseudolikelihood estimators, but these were not believed to be feasible due to the complexity of the data. Incidents that are duplicated across screener questions would pose the greatest difficulty.

[83] Imputation is a commonly used remedy for missing data, which involves filling in the missing values with allowable values for the variable in question. Many imputation methods have been devised (for an overview see Seastrom *et al* 2002).

Imputation review

In 2014 the NZCASS imputation process was reviewed in order to:

- assess the feasibility of including new imputation items for analysis and reporting, and if feasible, include these
- ensure the process was robust and fit for purpose
- ensure the imputation code was transparent and understandable to statistical experts outside the primary statistical provider (Statistical Insights)
- identify areas for improvement that could be either actioned as part of the 2014 project or recommended to future project teams.

As part of the 2014 imputation process, Independent quality assurance (IQA) was sought from experts at the University of Auckland for both:

- methodological changes
- updates made to the 2006, 2009 and 2014 imputation code.

During the 2014 analysis stage, some issues were identified which prompted further investigation into the imputation process and a second round of IQA. During this second round of investigation a number of errors and methodological issues with the 2006 and 2009 imputation code were identified. These errors and issues were reviewed, resolved and checked in order to ensure revised 2006 and 2009 estimates were true and correct.

Changes made in 2014

It should be noted that all changes, updates and corrections were made to all 3 iterations of the NZCASS (to ensure comparable statistics were produced) and quality assured by external experts.

Table 11.2: Changes made to 2014 imputation

Change	Description	Rationale	Notes
New 'Interpersonal violence' imputation items	Around 40 new imputation items were created to enable analysis of a range of relationship and violence groupings.	Users needed estimates by relationship type and type of violence in order to more accurately assess intimate partner and family violence issues.	
Treatment of double offence codes	Decision rules reviewed and updated when burglary is double coded with selected offences (ie when only burglary is included vs both offences).	Improve alignment between the NZCASS imputation process and Police practice.	Only affected vehicle offences.
Move from 10 to 100 imputations	The NZCASS has traditionally imputed for missing incident data with 10 values for each missing value. This was increased to 100 imputations in 2014.	Recent research suggested that 10 imputations would not be sufficient for some offence types, due to the amount of missing data in the NZCASS. This increase in the number of imputations decreased the amount of variance and substantially improved the power of significance tests.	
Separation problem/ Bayesian regression mode	Because sexual offences in SC3 had so few out-of-scope offences, this was causing the maximum likelihood regression model used as part of the imputation process to produce extreme values with large variability. This resulted in more out-of-s cope items than there should be ('separation problem').	In order to resolve the separation problem, a new Bayesian regression model was incorporated into the imputation process.	Options for resolving this issue along with recommendations were written up and reviewed by the ministry's IQA provider in the statistics department of Auckland University. Confirmation that the recommended option (the Bayesian model) was the most sensible methodologically and fit for purpose within the context of the NZCASS imputation process was gained prior to implementation.
Error corrections	A number of errors in the 2006 and 2009 imputation process/code were identified during 2014.	N/A	All errors identified were corrected and reviewed. The cumulative effect of these corrections, along with the inclusion of the Bayesian regression model, was that estimates for 2006 and 2009 increased from what was previously published.

Increasing the number of imputations

Previous NZCASS iterations (from 2001 on) have multiply imputed for missing incident data, with 10 values being imputed for each missing value (ie m = 10). In 2014 this decision was reviewed and an increase was agreed as desirable. Although the researchers who originally developed multiple imputation advised that 3–5 imputations would be enough in most situations, recent research has shown that values of m this small can substantially reduce the power of significance tests (Graham *et al* 2007). This is especially true when the fraction of missing information is moderately high, as in the NZCASS. Victim forms were missing for more than four- fifths of incidents overall in the 2014 dataset (46% for CAPI incidents and 96% for CASI incidents).

In the simple imputation problem studied by Graham *et al* (2007), for m = 10, power was depleted by around 10% compared to full information maximum likelihood when 70% of the information was missing, and by 15% when 90% of information was missing. Increasing the number of imputations reduced the loss of power, to 2.8% for m = 40 and to 1.1% for m = 100. To put these figures in context, a 15% drop in power would be equivalent to reducing the effective sample size from around 3000 to 2300 (for a 2-sided test with 95% confidence for the difference between 2 binomial proportions, 45% and 50%, in equally sized subgroups) – that is, by almost a quarter – whereas a 1.1% drop in power equates to reducing the effective sample size only to 2940.

As such, it was recommended that the number of imputations m increase from 10 to 100 for the 2014 NZCASS (and retrospectively for 2006 and 2009). Since the NZCASS has a substantial proportion of missing data (especially for CASI incidents), this will provide significance tests with substantially more power.

It was also noted that if needed, there is also the option of omitting a proportion of the imputed values for analysis and only using 10.

Imputing the number of incidents

Each screener question consisted of 2 parts:

1. ‘Since 1st January 2008, has anyone [done this to you]?’
2. then (if the answer to the first part was ‘Yes’) ‘How many times?’

The responses to the 2 parts were stored in separate variables for the CASI screener questions, but were combined into a single variable for the main screener questions (except Q42 and Q43).

The number of incidents was zero if the respondent reported that nothing of that nature had happened to them, but was missing if the respondent said ‘Don’t know/Can’t remember’ or ‘Don’t wish to answer’ at either part of the screener question. Table 11.3 shows the frequencies of these responses.

A value of 1 was imputed when the number of incidents was missing. In other words, it was assumed that the respondent reported being a victim of just 1 incident. This is likely to be a poor assumption in many of these cases, but it was hoped that this provided a reasonable compromise between over-counting for those who were not victims and undercounting for those who really experienced more than 1 such incident. However, it was suspected that most of these responses would be from victims, since non-victims would presumably not have trouble remembering the answer and might

have less reason to be averse to answering the question.[84] This suggests that the approach used here probably underestimates the true level of victimisation, and that other common imputation methods (such as an unrestricted random hot-deck) would also suffer from a similar problem.[85] Hot-deck imputation would also introduce more variability. The primary reasons for imputing a value of 1 in the 2014 NZCASS were that this approach was used in all previous surveys from 2001, and consistency was to be maintained.

The number of incidents was missing in 25 places for the CAPI screener questions, coming from 25 different respondents. The missing values were not uniformly distributed across the 15 screener questions; 60% of them (15) affected Q31 (attempted break-ins), accounting for 6% of the people who said 'Yes' to this screener question. Missing information was more common in the CASI section, with missing values accounting for 42% of the non-negative responses. Missing information was most prevalent at the sexual victimisation screener questions, where (except for the 'distressing sexual touching' question) almost 3 times as many respondents failed to provide information as provided a specific positive number of incidents. Even for the 'distressing sexual touching' question, missing information was as likely as complete positive information. These rates of missing data recorded at the sexual screener questions are substantially higher than those observed in the 2009 NZCASS.

If mean imputation or random hot-deck imputation was used instead of imputing a value of 1, still assuming that all the missing responses were from victims (and restricting the mean or the donor pool to victims accordingly), this could almost quadruple the estimated incidence of rape. The estimated prevalence would also be somewhat higher (because the higher number of incidents means it is more likely one will still count after imputation of dates and relevance), though prevalence would not increase as much as incidence. The estimated incidence and prevalence of all offences would increase slightly. Hot-deck imputation would be essentially unbiased if all the respondents who do not know or refuse to state the number of incidents they experienced (or whether they experienced any at all) were victims, and they experienced similar numbers as respondents who did acknowledge being victims. Mean imputation relies on similar assumptions.[86] At least the first of these assumptions seems unlikely to hold, however, and we have no strong reason to believe that the biases caused by these 2 aspects of the issue would necessarily cancel out.

Table 11.3 shows the numbers of respondents giving definite answers at each screener question, as well as the numbers not giving a response for various reasons, and the average number of incidents reported among those giving a positive number. Structural zeroes are denoted with a dash.

[84] However, non-victims might have refused to answer the CASI screener questions because they felt the topic was too personal in nature, despite not being victims themselves. Refusals were the dominant form of non -response for the sexual screener questions.

[85] This assumes the donor pool would consist of all respondents with complete data for that screener question. Another possibility is to restrict the donor pool to those reporting some incidents at that screener, which would probably then err in the other direction.

[86] Mean imputation would induce some technical bias in prevalence estimates, even if its underlying assumptions hold, due to the non-linear effect of other imputation steps on these estimates.

Table 11.3: A summary of responses at the screener questions

Screener	Zero	More than zero	Don't know/ Can't remember	Refused	Not asked	Average non-zero incident count
Q28	6,231	118	0	-	594	1.31
Q29	5,991	358	0	-	594	1.25
Q30	5,876	473	0	-	594	1.43
Q31	6,681	247	15	-	-	1.74
Q32	6,686	256	1	-	-	1.48
Q34	6,488	451	4	-	-	1.47
Q35	6,751	189	3	-	-	2.34
Q35_416	6,642	301	0	-	-	1.82
Q36	6,812	131	0	-	-	1.47
Q37	6,659	284	0	-	-	2.39
Q38	6,885	58	0	-	-	1.48
Q39	6,883	60	0	-	-	1.72
Q40	6,878	65	0	-	-	1.35
Q41	6,765	178	0	-	-	1.55
Q42/Q43	6,769	172	2	-	-	3.40
Q167_419/420	4,261	90	35	40	2517	22.64
Q167_421/422	4,287	105	11	23	2517	39.24
Q167_423/424	4,307	83	19	17	2517	12.48
Q167_425/426	4,333	60	18	15	2517	18.23
Q227/228	6,548	176	51	37	131	20.18
Q229/230	6,534	198	42	38	131	15.63
Q231/232	6,622	103	53	34	131	18.40
Q233/234	6,656	86	37	33	131	24.23
Q287_433/434	6,743	19	13	37	131	14.89
Q287_435/436	6,735	21	16	40	131	10.67
Q287_437/438	6,674	69	23	46	131	6.13
Q287_439/440	6,741	17	14	40	131	10.71

Another cause of missing information for the CASI screener questions was the refusal of 1.9% of respondents to complete the CASI section. No overt imputation has been conducted to correct for this – that is, it is effectively assumed that these people experienced no offences of the types covered by the CASI screener questions. This will have led to underestimation of the true victimisation rates for these offence types, although the bias will not have been large due to (1) the small level of CASI non-response, and (2) its skew towards older respondents. Previous analysis of the age profile of the 2009 NZCASS CASI non-response and of confrontational crime from the CASI

screener questions suggested the incidence of confrontational crime among CASI non-respondents might be around 70% as high as among the rest of the sample. CASI non-respondents in the 2014 NZCASS tended to be younger than in 2009, which suggests that while the incidence of crime reported primarily at the CASI screener questions may be slightly understated due to CASI non-response, this understatement may well be less than 1.3%. CASI non-response bias thus appears to be a minor issue.

Date of imputation

For each incident without a victim form from the CAPI screener questions, the calendar year in which the incident occurred was imputed randomly assuming that it had an equal chance of occurring on each day between 1 January 2013 and the interview date. That is, the year each incident occurred was imputed as being in 2013 with probability equal to 365 divided by the number of days between 1 January 2013 and the interview date. This was done independently across incidents and for each of the 10 imputations conducted per incident. This is the same method as used in the 1996, 2001, 2006 and 2009 surveys.

For CASI incidents, the same method was used, except when the incident with date information from that part of the CASI section occurred during 2013. Since that incident is the last incident in that part of the CASI section to have occurred, all the others are then imputed as occurring during 2013. The same method has been applied to data from the CASI sections of the 2006 and 2009 surveys. The assumption of even spread may not be ideal when the last incident occurred in 2013, because knowing this provides some additional information about when the other incidents are likely to have occurred.

The assumption of even spread also does not account for recall bias. An investigation of the known incident dates in 2006 (described in Appendix A6 to the 2006 NZCASS Technical Report) suggested that this is likely to have had a substantial effect on the victimisation risk estimates from that survey, and even stronger effects in the previous surveys.[87] However, no easy method of correcting for this has been apparent, and imputing using the empirical date spread would actually make things worse.

The interviewing dates in 2014 were similar to those in 2006 and 2009, so any recall effect will probably have affected these 3 surveys to much the same extent. Comparisons over this time period should therefore be relatively unaffected.

[87] Similar patterns of bias were observed in the 2005 Irish International Crime Survey (van Dijk *et al* 2007: 9 –11).

Relevance imputation

Different types of offences have widely varying relevance rates (and varying proportions of missing data). These are shown in Table 11.4, broken down by source question (ie the screener question at which that incident was enumerated). Here 'relevant' means that the incident had been assigned an offence code other than 85, 86 or 87, involved the type of offence covered by the NZCASS in the section of the questionnaire where it was reported (as recorded at Q47.373, Q182, Q235 or Q318), occurred in New Zealand since 1 January 2013 (confirmed at Q44), and was not the same incident as covered by a preceding victim form (ascertained at Q46.403 and Q46.448).

Table 11.4: Missing forms and relevance rates by screener question

Source question[88]	Description	% of incidents without victim forms	Relevance rate (for completed victim forms)
Q28	Vehicle theft	23	82
Q29	Theft from vehicle	22	81
Q30	Damage to vehicle	27	70
Q31	Attempt to break in	47	65
Q32	Burglary	35	83
Q34	Theft from property	31	83
Q35	Theft from inside home	53	82
Q35.416[89]	Household damage	34	81
Q40	Theft from person	28	83
Q41	Other theft	32	56
Q36	Assault	31	82
Q37	Threat of assault	51	62
Q38	Other damage	36	60
Q39	Threat of damage	57	55
Q43	Other incidents (CAPI section)	64	41
Q167.420	Assault (by current partner)	98	85
Q167.422	Threat of assault (by current partner)		
Q167.424	Damage (by current partner)		
Q167.426	Threat of damage (by current partner)		
Q228	Assault (by person well known)	95	75
Q230	Threats (by person well known)		

[88] The source question is not recorded for CASI victim forms, so the rates shown apply to a whole part of the CASI section (ie 4 screener questions combined).

[89] Questionnaire numbering: Q416 followed Q35; it is referred to as Q35.416 to indicate this. The survey documentation follows this protocol in most cases where question numbers did not reflect the interview order.

Source question[90]	Description	% of incidents without victim forms	Relevance rate (for completed victim forms)
Q232	Damage (by person well known)		
Q234	Threat of damage (by person well known)		
Q287.434	Forced sexual intercourse	90	98
Q287.436	Attempted forced sexual intercourse		
Q287.438	Distressing sexual touching		
Q287.440	Other sexual violence, incl. threats		

The relevance rates for the CAPI screener questions vary from 41% to 83%, so the screener question was expected to be a useful predictor of relevance status. Relevance status was imputed for incidents from the CAPI section using a Bayesian logistic regression model[91] that included the screener question, household composition, household size, tenure/landlord, gender, age group (15–24, 25–39, 40–59, and 60 years or more), marital status, employment status, ethnicity (European, Māori, and Other), urbanisation and the NZDep13 score.[92] Details of the model are shown in Appendix F. This model was used to multiply impute relevance status for incidents from the CAPI screener questions without victim forms, by generating parameter values from their posterior distribution for each of the 100 imputations, using these to predict the probability that each incident would be relevant, then randomly generating relevance status using these predicted probabilities.

A very similar process was followed for CASI incidents. The only differences relate to the screener questions. First, the screener questions in the CASI section are different from those in the CAPI section. Also, the source screener question for the last incident is not collected directly. It was instead imputed randomly within each part of the CASI section with probability proportional to the number of incidents reported at each screener question. Details of the imputation model for CASI incidents are shown in Appendix F.

[90] The source question is not recorded for CASI victim forms, so the rates shown apply to a whole part of the CASI section (ie 4 screener questions combined).

[91] As mentioned earlier, Bayesian logistic regression was adopted in the 2014 NZCASS due to separation problems. It also has the advantage of yielding useful results with a more complex imputation model, avoiding the model selection stage needed in previous surveys. Data from the 2006 and 2009 surveys was reimputed using the 2014 approach to enable better comparisons over time.

[92] Respondents with missing values for any of the predictor variables were omitted from the model.

Offence code and offender knowledge/relationship imputation

Offence codes were imputed using a hot-deck imputation method (the approximate Bayesian bootstrap; Rubin 1987: 124), with imputation classes defined by source screener question. In other words, the donor pool used to impute the offence codes for incidents from a given screener question was a with-replacement sample of incidents from that same screener question (the imputed screener question was used for CASI incidents). This technique reproduces the distribution of offence codes from each screener question, on average, and aims to incorporate an appropriate degree of variability into the multiple imputations. The same method was used in the 2006 and 2009 surveys.

The same technique, and in fact the same donors, were used to impute the victim's knowledge of and relationship to the offender (or offenders). This ensured that statistical associations between offender knowledge/relationship and offence codes were not attenuated. In previous iterations of the NZCASS, the victim's knowledge of and relationship to the offender(s) had not been imputed, but this became necessary in the 2014 NZCASS to produce the desired measures of interpersonal violence. The same technique was applied to data from the 2009 and 2006 surveys to enable comparisons of these measures over time.

See Appendix F for further details on the imputation of interpersonal violence, in particular how the weighting and imputation procedures correctly handle the counting of ex-partner violence.

Duplication of incidents

Respondents could potentially duplicate the reporting of incidents within the NZCASS questionnaire by:

1. reporting the same incident multiple times within the CAPI victim forms
2. reporting the same incident across the 3 CASI victim forms
3. reporting the same incident in both the CAPI and CASI sections

The NZCASS questionnaire is designed to minimise the chance of this occurring. The screener questions are intended to gather the number of incidents experienced by each respondent, without double counting. To make this clear to the respondent, all the screener questions in both CAPI and the first 2 parts of CASI (apart from the first screener question in each section of the questionnaire) incorporate phrases such as 'Apart from this' or 'Leaving aside anything already mentioned'.

(1) Duplication within CAPI

In 2006 a new set of questions was added to the CAPI victim forms to establish how well the instructions worked to not repeat incidents already mentioned. Specifically, these questions asked whether the current incident was actually the same as in one of the previous victim forms (if the incidents came from different screener questions). If so, the rest of the victim form was skipped. These duplication questions were retained in 2009 and 2014.

Because some incidents were not selected for victim forms, the duplication questions only provided partial information on the degree of overlap or duplication between screener counts. To be specific, they only detected duplication between the 3 (or fewer) incidents for which victim forms were completed.

For example, if only 3 incidents were reported in the CAPI screener questions (call them A, B and C), the new questions would give complete information about the 3 possible duplications AB, AC and BC. If 1 extra incident D was reported (at a separate screener question), no information would be gathered about the 3 extra possible duplications AD, BD and CD. If yet another incident E was reported, there would be no information about its possible duplications AE, BE, CE and DE, and so forth. The number of unobserved duplications can thus be quite large for heavily victimised respondents.

The unobserved duplications were imputed for other incidents by estimating the rate of duplication per potential clash, and independently simulating duplications randomly for each unobserved potential clash, using the estimated duplication rate.

Table 11.5: Estimated duplicated incidents within CAPI

	2006	2009	2014
Rate of duplication per potential clash	3.79%	5.45%	4.26%
Percentage of incidents from CAPI screener questions with no victim form estimated to be duplicates	9.0%	11.2%	7.5%
Observed number of duplicated and triplicated incidents	95 duplicates 1 triplicate	182 duplicates 5 triplicates	86 duplicates 6 triplicates

The rate of duplication per potential clash was estimated separately for incidents with 1 and 2 potential clashes, and these rates were averaged to derive an overall estimated duplication rate per potential clash.

In 2014 approximately 7.5% of incidents from the CAPI screener questions with no victim form were projected to be duplicates. This is higher than the rate of duplication per potential clash (4.26%), because most of the incidents had many potential clashes. The projected duplicates were excluded from the calculation of victimisation rates to help avoid over-reporting through failure to follow the 'apart from…' instructions.

(2) Duplication within CASI

No such questions were added to the CASI section because victim form questions were only asked for 1 incident from each part. Because of this, a similar assessment and adjustment of duplication within each part of the CASI section is not possible.

It would be possible to add questions to establish whether the CASI victim forms referred to the same incident as a victim form from an earlier part of the CASI section, although in theory the chance of this happening is small due to the separate topics of the 3 CASI parts. In particular:

- **SC1 (PARTNER) AND SC2 (WELL KNOWN):** It is very unlikely for there to be duplication between SC1 and SC2 since by nature of the collection the relationship to offender differs. For a small number of cases with multiple offenders, the incident could include both a partner and other people well known. However, relationship to offender is not collected in SC1 (as all partners), and partner is not a valid option in SC2, and consequently there is no information to assess potential duplicates.
- **SC3 (SEXUAL) WITH SC1 (PARTNER) OR SC2 (WELL KNOWN**): If the offender of the sexual offence was a partner/person well known, then in theory the same incident could be reported in both SC3 and SC1/2. However from an analysis of offence codes and relationship to offender, there were no records at all in 2014 that could be potential duplicates between SC3 and SC1/2, and in 2009 this was very small at 0 and 2 respectively. Even if these 2 records were assessed to be actual duplicates, the issue is very minor to consider further.

(3) Duplication between CAPI and CASI

The new interpersonal violence framework introduced in 2014 meant that information about relationship to the offender and offences was gathered together from throughout the questionnaire (both CAPI and CASI victim forms), rather than just from 1 section (as was the case previously with the 'confrontational crime' framework).

For example, the new items 'current partner offences' are obtained from across the NZCASS questionnaire where the relationship to offender was a current partner and if the offence codes meet the definition of interpersonal violence. Previously (for confrontational crime items) only incidents that were reported in CASI victim form 1 (violence by a current partner) were included. The new approach was considered more appropriate since many incidents involving current partners were reported in CAPI victim forms and CASI victim forms (eg damage to personal property by a current partner).

This new framework meant that duplications within CASI and CAPI victim forms could have a larger impact than in previous NZCASS iterations, with the risk of the new approach over-counting if respondents reported incidents in multiple parts of the questionnaire.

Cases of potential duplication

To address this point, duplication was assessed between the 3 CAPI victim forms and each of the 3 CASI victim forms (separate potential duplication within CAPI and CASI has been addressed above).

Potential duplicates were defined according to the rules set out in Table 11.6.

Table 11.6: Rules defining potential duplicates between CAPI and CASI

Topic	Coding rules
Violence by a partner	Primary or secondary offence code = 6, 7, 8, 9, 21, 29 AND Q65 or Q74 = 1, 3 AND a reported SC1 incident (not an 80s code)
Violence by people well known	Primary or secondary offence code = 6, 7, 8, 9, 21, 29 AND Q65 or Q74 not equal to 1, 3 AND a reported SC2 incident (not an 80s code)
Sexual offences	Primary or secondary offence code = 1, 2, 4 AND a reported SC3 incident (not an 80s code)

Notes:

1. These offence codes were used due to the definition of confrontational offences used in 2009 and 2006. Following this research of potential duplication, it was proposed to change confrontational crime to interpersonal violence. The frequencies would change slightly if the interpersonal violence offence codes were used instead of the confrontational crime offence codes, but as the changes are minor then the assessment would be very similar.

2. Incidents with an 80s code (85, 86, 87) were ignored as none of these codes are included in analysis.

Table 11.7 shows the initial number of potential duplicates between the CAPI victim forms and the 3 CASI victim forms. This presents there were a smaller number of potential duplicates in 2014, particularly compared to 2009 (reasons for this are provided below).

Table 11.7: Frequencies of potential duplicates between CAPI and CASI, by year

Topic	Victim form duplication	2006 frequency	2009 frequency	2014 frequency
Violence by a partner	CAPI victim forms and SC1	10	16	3
Violence by people well known	CAPI victim forms and SC2	5	38	1
Sexual offences	CAPI victim forms and SC3	2	2	2
Total		**17**	**56**	**6**

Independent coding advice on duplication

The records for potential duplicates in 2006 and 2009[93] were sent to the NZCASS legal coding expert at Victoria University for review regarding the certainty of duplication from a review of the text field descriptions and other fields used during the coding process. The findings were as follows:

Table 11.8: Assessment of duplication from potential records between CAPI and CASI

Topic	2006	2009
Duplicate	1	7
Could be duplicate (not enough information)	4	7
Not duplicate	12	42
Total	17	56

Of those classified as 'duplicate' in Table 11.8, the coding expert gave a rating from 1 to 5 where 1 = not at all confident to 5 = certain. Of the 7 incidents assessed as duplicated in 2009, only 1 respondent's incidents were assessed as a 5, with the remaining being 2, 3 or 4. In 2006 the level of certainty for the 1 instance was a 4.

For the not-duplicates, the expert provided reasons why they were not duplicated incidents including being a series of similar offences, different offence types, different offenders, different location of incident, and different incident descriptions.

Given this, while the issue of duplication can exist when selecting violent interpersonal offences across the whole questionnaire instead of within sections, it is insignificant in scale. On balance, the benefits of moving towards the method of capturing incidents across the questionnaire rather than within-section far outweighs the issue of over-counting due to duplication of incidents.

The advice was to only discard duplicated incidents where it was certain they were duplicates. From this can be seen that there were only a very small number of instances that were assessed as duplicates, and even these were not deemed definitive.

[93] 2014 potential duplicates were not sent to the coding expert as the assessment was done in May 2014 prior to the 2014 final dataset being available.

In addition to the assessment of potential duplicates, the coding expert also provided advice on the following 2 points on coding practices:

1. In cases where the coders assessed the incidents as duplications they coded the second incident as '86: Not an offence'.[94] The offence was coded in the more appropriate section (eg if sexual offence then given code in SC3 and 86 in the other section). This was done for each survey year.
2. The fact that there were substantially more potential duplicates in 2009 was queried, and it was suspected that this was largely due to the coding interface/system used. In 2009 the coders were presented with the CAPI victim forms first, and then each CASI victim form separately. All information was provided in an Excel workbook, which was cumbersome. As the coder did not see all victim forms for each respondent together, it was difficult for them to assess potential duplication and relied on the checker to pick this. In 2014 an online and specifically tailored coding interface was developed (Appendix D). Each respondent's victim forms were able to be seen together; consequently, duplication was more recognisable to the coder.

These 2 points confirm that duplication is taken into consideration during coding practices, and support the decision to conduct no further adjustments during analysis.

[94] The specific instructions from the Offence Coding Manual were:
'Self Completion [SC] and Victim Form [VF] duplicates:
- If you are sure that both SC and VF forms relate to the same incident, code 86 should be applied to one of the forms to avoid duplication and double counting. Apply code 86 to the VF if the offence was committed by a partner or someone well known to the respondent. Apply code 86 to the SC if the offence wasn't committed by a partner or someone well known.
- If there is a SC and VF reports of offences by people well known that relate to different incidents (not duplicates) then all incidents should be coded but noted as uncertain for review by supervisors/auditors.
- If an incident by someone well known is reported in a VF and there is no SC form completed by the respondent, the VF should be coded as usual.'

Heavy victimisation cut-off

After imputation, a cut-off was applied to improve the reliability of the estimated rates. Using the imputed values, all out-of-scope or duplicated incidents, and those that did not occur in New Zealand during 2013, were excluded from the dataset. The number of remaining offences from the CAPI section was not allowed to exceed 30 for any respondent. Any further offences above this value were not included in the victimisation estimates. The specific incidents to be deleted were selected randomly, and this was done separately for each of the 100 imputations. The same cut-off procedure was applied independently to incidents from the CASI section.

Averaged across imputations, this cut-off ruled out 1.1% of incidents from the CAPI section that would otherwise have been counted. These came from 15 respondents (ie 0.2% of all respondents). The cut-off had a much greater effect on CASI incidents, ruling out 58.2% of these on average (again restricting consideration just to those incidents that would otherwise have been counted). These came from just 46 respondents (0.7% of all CASI respondents[95]).

The same cut-off value and method was applied in 2006 and 2009. The percentage of CAPI incidents that were dropped in the 2014 NZCASS was lower than in the 2009 survey, which in turn was lower than in 2006, but the numbers of respondents affected were similar. In contrast, the percentage of CASI incidents that were dropped in 2014 was higher than in previous years.

No cut-off was applied in 2001. The introduction of the cut-off in 2006 was prompted by the easing of controls on how many incidents could be reported at each screener question. The selection of the cut-off value was partly guided by an examination of the trade-off between the estimated reductions in variance that would be achieved against the potential bias each cut-off value might introduce.

The impact of this cut-off has been assessed in previous years as having minor impacts on the standard NZCASS measures, including incidence rates. However the heavy victimisation cut-off particularly affects distributional measures of crime, for example how many offences are experienced by how many people and the Gini coefficient. Such statistics therefore need to be interpreted with caution, since concentration will tend to be underestimated due to this heavy victimisation cut-off.

[95] Based on CASI respondents, since it was impossible for people who refused to answer the CASI questionnaire to exceed the CASI cut-off.

Review and quality assurance process

The review and quality assurance process was undertaken in 2 key stages (Table 11.9).

Table 11.9: Imputation review and quality assurance process

Stage	Type of review	Description	Process
1	General review	To understand what areas of the imputation process need to be updated as part of the 2014 NZCASS.	1. Review of documentation, past decisions and leanings from project teams. 2. Recommendation on which changes (if any) are feasible for the 2014 NZCASS along with rationale.
	Methodological review of new imputation items	To assess which imputation method is best suited in order to incorporate the new interpersonal violence items.	1. Proposal developed by statistical services provider. 2. Proposal reviewed and commented on by methodological expert from the statistics department at Auckland University. 3. Recommendation finalised by statistical services provider. 4. Recommendation confirmed by external IQA and IQA report provided to the ministry.
	Code updated for each year	Starting with 2014 and working backwards in reverse order, each year's imputation code was reviewed and updated in order to incorporate the agreed changes.	1. Code updated by statistical services provider. 2. Code reviewed by statistical coding expert from the statistics department at Auckland University. 3. Collaborative, real-time discussions were undertaken between statistical providers in order to refine code. 4. IQA report provided to the ministry on the review of each year's code.
	Data delivery	Imputed data delivered to Ministry.	Internal Ministry checks undertaken.

Stage	Type of review	Description	Process
2	Issues review	While analysing data as part of reporting processes, the ministry found several anomalies which were rooted in the imputation process developed/used in previous years (2006 & 2009)[96].	1. Statistical Insights reviewed the imputation process to identify the areas which were most likely to have caused the anomalies. 2. Quality assurance expert from the statistics department at Auckland University was then contracted to undertake a line-by-line review of the old and revised code according a prioritised list of areas. 3. Quality assurance expert logged discrepancies in the code and sent these to statistical services provider for further investigation rescale and impact. 4. Statistical provider investigated selected areas of the code in parallel with quality assurance expert's review. 5. Ongoing project meetings were held with external experts and project team to discuss discrepancies, their possible impact, prioritise investigation and document progress. 6. An 'Imputation issues log' was compiled, updated by project team members to form part of the project record. 7. All resolved issues were checked. 8. A quality assurance report along with recommendations was compiled and provided by the external quality assurance expert.

[96] It should be noted that issues did not originate due to changes in the imputation process for NZCASS 2014 but were present as errors in 2006 and 2009.

12. Reporting analysis

How variable are results

Sampling error

The NZCASS is a sample survey. This means that a sample of areas (meshblocks), households and people are selected from the New Zealand population using a set process. Sampling error arises because only a small part of the New Zealand population is surveyed, rather than the entire New Zealand population (census). Because of this, the results (estimates) of the survey might be different from the figures for the entire New Zealand population. The size of the sampling error depends on the sample size, the size and nature of the estimate, and the design of the survey.

Relative standard error

In the NZCASS, the relative standard error (RSE) is used to measure the variability of count estimates or means (rather than percentages). The RSE is reported as a percentage of the estimate: RSE = standard error of the estimate/estimate.

In NZCASS reporting:

- estimates with an RSE between 20% and 50% are considered high and are flagged with a hash symbol (#); we do not recommend using flagged estimates for official reporting (eg ministerial reporting)
- estimates with an RSE over 50% are suppressed as they are considered too unreliable.

The NZCASS relative standard error is similar to the relative sampling error used by Statistics NZ, but is not the same.

Margin of error

In the NZCASS, the margin of error (MoE) is used to measure the reliability of percentage estimates (rather than counts or means) and to calculate confidence intervals. NZCASS reporting uses the 95% Mo E and this is calculated as the t-value (approximately 1.96) multiplied by the standard error: MoE = t-value × standard error of estimate.

In NZCASS reporting:

- estimates with an MoE between 10% and 20% are considered high and should be viewed with caution
- estimates with an MoE over 20% are suppressed as they are considered too unreliable for general use.

Confidence intervals

Confidence intervals (CIs) are also used to measure of an estimate's reliability. A CI expresses the sampling error as a range of values between which the 'real' population value is estimated to lie. The 95% CI is used in NZCASS reporting, and is calculated as the estimate plus or minus the Mo E: CI = estimate ± MoE of the estimate.

Statistical significance

All significance testing in the 2014 NZCASS was done at the 95% confidence level, unless otherwise stated. See the 'Analysis Methods' section for how significance tests were conducted.

How to calculate different measures of variability

The NZCASS uses a confidence level of 95%. If you need a different confidence level, you can calculate them with the following approximations:

- 90% confidence level: multiply the MoE by a factor of 1.645/1.96
- 99% confidence level: multiply the MoE by a factor of 2.576/1.96

We have used RSE to measure the reliability of count estimates and means, and MoE to measure the reliability of percentages. If you need to use a different measure, you can convert them using the following approximations:

RSE to MoE: MoE ≈ (RSE × estimate)/100 × 1.96

MoE to RSE: RSE ≈ (Mo E × 100)/(1.96 × estimate)

If you need different statistical significance tests, you can get a broad approximation by examining whether the CIs overlap. If the CIs do not overlap, then it can be said the difference is statistically significant. If the CIs do overlap, then it is likely (but not definite) that the difference is not statistically significant.

Non-sampling error

The variability of results (sampling error) discussed in this section should not be confused with non-sampling error, which also contributes to survey error. Non-sampling error includes inaccuracies that can arise through reporting to interviewers, respondents' memory issues or fabricating information, errors in coding and processing, non-response bias and inadequate sampling frames. While inaccuracies of this kind may happen in any survey, strict quality processes have been undertaken in the NZCASS to minimise this type of error. While sampling error can be quantified, non-sampling error cannot.

Analysis Methods

What statistics were used in NZCASS reporting

The three types of statistics used in NZCASS reporting are:

- **Percentages** – for example 24% of adults were a victim of crime once or more in 2013.
- **Averages** – for example on average 69 offences were experienced for every 100 adults in 2013 (incidence rates). Incidence rates are a type of average, and averages are also used for the incident mean seriousness score.
- **Totals** – for example the estimated total number of burglary offences in 2013 was 203,000. Totals are mainly used in crime rate tables.

These statistics are always weighted and in the data tables given with their sampling error. While we used standard methods to analyse these statistics, NZCASS has some complexities that we explain in this section:

- the use of replicate weights to estimate sampling error
- the use of imputed data
- the tests used to assess statistical significance.

How sampling error was calculated

We used a replicate weight technique to estimate sampling error for NZCASS. These methods are well established for use in surveys with complex sample designs and weighting procedures.

The method used was the delete-a-group Jackknife method, where 100 replicate weights were calculated to accompany the main weight. See Chapter 10 for how we calculated the main weights. We calculated replicate weights by dividing the main sample into 100 groups, and leaving out one group from the sample at a time to form each replicate. We re-ran the weighting process for each of the 100 replicate groups, to produce 100 replicate weights.

We then used these 100 replicate weights to calculate the sampling error of the estimate. This is done by calculating the statistic of interest for the main weight and for each of the 100 replicate weights. The variance estimate for unimputed data in NZCASS 2014 is then calculated as:

Equation 1

$$\hat{v}(\hat{\theta}) = \frac{99}{100} \sum_{k=1}^{100} (\widehat{\theta_k} - \theta)^2$$

where θ is the statistic of interest calculated using the main survey weights, and θ_k is the same statistic calculated using the k^{th} set of replicate weights.

The standard error of the estimate is the square root of this variance estimate. This standard error is then used in the calculation of the relative standard error, the margin of error and confidence intervals. See earlier in this chapter for the definitions of these measures. For confidence intervals of unimputed data, the t-value has 99 degrees of freedom.

We give the sampling error for all NZCASS statistics in the NZCASS data tables.

How imputed data was analysed

Imputation is a statistical method to account for missing data – see Chapter 11. The imputation process assigns 100 imputations for each respondent, which means that the analysis of imputed data is more complex than the standard formula. We calculated estimates and variance estimates for

each of the 100 multiple imputation datasets. These statistics were then combined across the 100 imputations using Rubin's (1987) combining rules as follows:

The overall estimate is the average of estimates from each individual imputed dataset:

Equation 2

$$\hat{\theta} = \frac{1}{J}\sum_{j=1}^{J}\widehat{\theta_J}$$

where J is the number of imputations (100), and $\widehat{\theta J}$ is the estimate of the statistic for the J^{th} imputed dataset.

We calculate the variance estimate for statistics involving imputed data from both the within-imputation variance and the between-imputation variance.

Within-imputation variance:

Equation 3

$$\overline{U} = \frac{1}{J}\sum_{j=1}^{J}\widehat{v_j}\,(\theta)$$

$\hat{v}j(\widehat{\theta})$ is the variance estimate for the j^{th} imputed dataset, calculated using replicate weights, as specified in Equation 1.
Between-imputation variance:

Equation 4

$$B = \mathrm{var}\left[\widehat{\theta_j}\right],$$

where $\hat{\theta}_j$ is the estimate of the statistic for the j^{th} imputed dataset, and $var\left[\hat{\theta}_j\right]$ is the variance of the statistic's estimate across the j imputed datasets.

This is combined to yield the total variance estimate:

Equation 5

$$\hat{v}(\widehat{\theta}) = \overline{U} + \left(1 + \frac{1}{J}\right)B.$$

As we explained above, the standard error is the square root of this variance estimate. Confidence intervals are calculated as the estimate (θ) plus and minus the standard error multiplied by the t-value. For imputed data, the t-value has the following degrees of freedom for 100 imputations:

Equation 6

$$df = 99 \times \left(1 + \frac{100 \times \overline{U}}{101 \times B}\right)^2$$

How tests of statistical significance were conducted

When estimates are compared over time or between sub-groups within a survey year, the differences between the estimates are tested for statistical significance. We do this to determine if the difference is 'real' or simply caused by the survey's sampling error. We used different significance tests for tests over time and for tests between categories within a survey year. Statistical significance was tested at the 95% confidence level (unless otherwise stated).

Comparisons over time (between surveys)

Let $\hat{\theta}_1$= estimate for time 1 (eg 2009)

$\hat{\theta}_2$ = estimate for time 2 (eg 2014)

The difference is (y) = $\hat{\theta}_1 - \hat{\theta}_2$. The standard error of the difference is approximated by:

Equation 7

$$\mathrm{SE}(\hat{\theta}_1 - \hat{\theta}_2) = \sqrt{\mathrm{SE}(\hat{\theta}_1)^2 + \mathrm{SE}(\hat{\theta}_2)^2}$$

If the confidence interval for this difference contains 0, then the difference is considered not statistically significant.

Comparisons between sub-groups within one survey

We used different tests of significance for different statistics:

- t-tests were used to determine statistical differences in averages (incidence rates and means) and totals
- rate ratios were used to determine statistical differences in percentages (including prevalence rates).

t-test

Let $\hat{\theta}_A$= estimate for Group A

$\hat{\theta}_B$ = estimate for Group B

The difference is (y) = $\hat{\theta}_A - \hat{\theta}_B$. The variance of the differenc e is calculated using Equation 1 where $\hat{\theta}$ in this equation is the difference (y). Likewise, variance estimates for data involving imputation are given by Equation 5.

If the confidence interval contains 0, then the difference is not considered statistically significant.

Rate ratio test

Let $\hat{p}_A$= percentage for Group A

$\hat{p}_b$ = percentage for Group B

The rate ratio $\hat{r}_{A,B} = \hat{p}_A/\hat{p}_B$. The variance of the rate ratio is calculated in a log scale using Equation 1 where $\hat{\theta}$ in this equation is the log of the rate ratio, ie $\mathrm{In}\left(\hat{r}_{A,B}\right)$. Variance estimates for rate ratios for data involving imputation are given by Equation 5. The confidence interval is formed by taking the exponential of the lower and upper bounds calculated from the log of the rate ratio.

If the confidence interval contains 1, then the difference is not considered statistically significant.

Advanced statistical methods

Different types of analyses have been used to help us better summarise and understand different aspects of crime and victimisation. The 3 statistical methods of modelling used as part of the NZCASS and described in this section are:

- multiple standardisation
- logistic regression
- the Gini coefficient

This section of the Technical Manual presents the aims, methods and more detailed results to that presented in the Main Findings Report.

Multiple standardisation

The NZCASS main findings report and data tables show us that 32.9% of Māori had been victimised once or more in 2013, compared to 22.6% of the European population. This difference in victimisation is 10.3 percentage points.

Since Māori are over-represented in lower socio-economic groups and have a younger population, is victimisation really about being Māori or more to do with poverty or some other factor ? This raises the question: If these factors (such as socio-economic status) were standardised across ethnic groups, are Māori still more at risk of being victimised?

Factor (variable) selection
There are a range differences between the Māori and European populations. For example, a slightly higher proportion of Europeans live in major urban areas than Māori, but whether this is a strong difference relevant to victimisation needs to be tested to determine what factors (variables) to standardise by.

While many different factors could be considered, only 2 or 3 should be used due to sample size restrictions. This is supported by the Australian Institute of Health and Welfare (2011) principle that recommends a sample size of approximately 20–30 in each category for standardisation to be appropriate. This means that for the

NZCASS sample size, standardisation should be conducted on approximately 2 or 3 variables (depending on category numbers).

Age and NZDep2013 quintiles were selected as the factors to standardise by. These 2 factors were selected because:

- User feedback from previous iterations of NZCASS analysis raised that socio-economic status is a ‘confounding’ reason for the higher rates of Māori victimisation. This means that when analysing the victimisation differences between Māori and Europeans, the question remains whether the difference is due to socio-economic differences rather than ethnicity, since lower socio-economic groups are more likely to be victimised and proportionately more Māori have lower socio-economic status. NZDep2013 was selected as the socio-economic measure since it is a multi - dimensional measure.
- In addition to socio-economic status, age was selected due to the age profile differences in the Māori and European populations. The 2013 Census shows the large age differences in the Māori and European populations; for example, Māori have a median age of 24 years, and Europeans have a median age of 41 years (Statistics NZ 2013).
- This was further supported by the data through using a decision tree analysis to determine the prominent drivers of victimisation (similar approach to the Statistics NZ (1998) paper). Once this analysis was performed, the 2 variables of age and deprivation emerged as important.

This analysis does not attempt to control for all differences between Māori and Europeans, but rather to consider some of the main differences to then assess the size of the victimisation risk gap once these factors are controlled for.

Structural differences of age and deprivation

Viewed graphically we can see the profile differences between Māori and Europeans for age (Figure 12.1) and deprivation (Figure 12.2). In comparison to Europeans, Māori are a younger population and proportionally more Māori live in areas of higher deprivation.

Figure 12.1: Māori and European population: profile of age groups (NZCASS 2014)

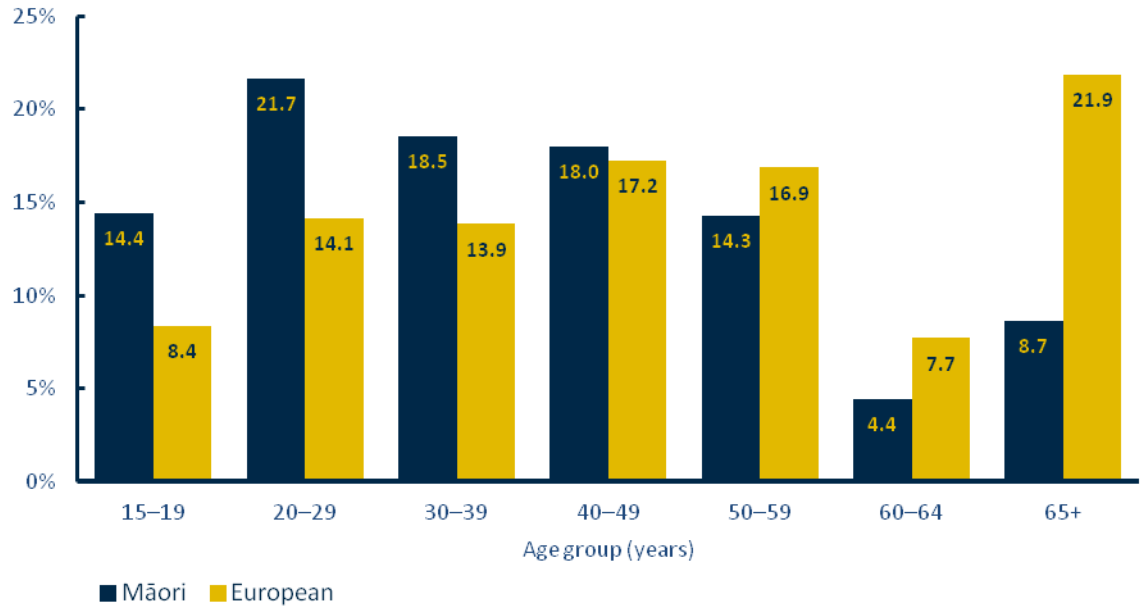
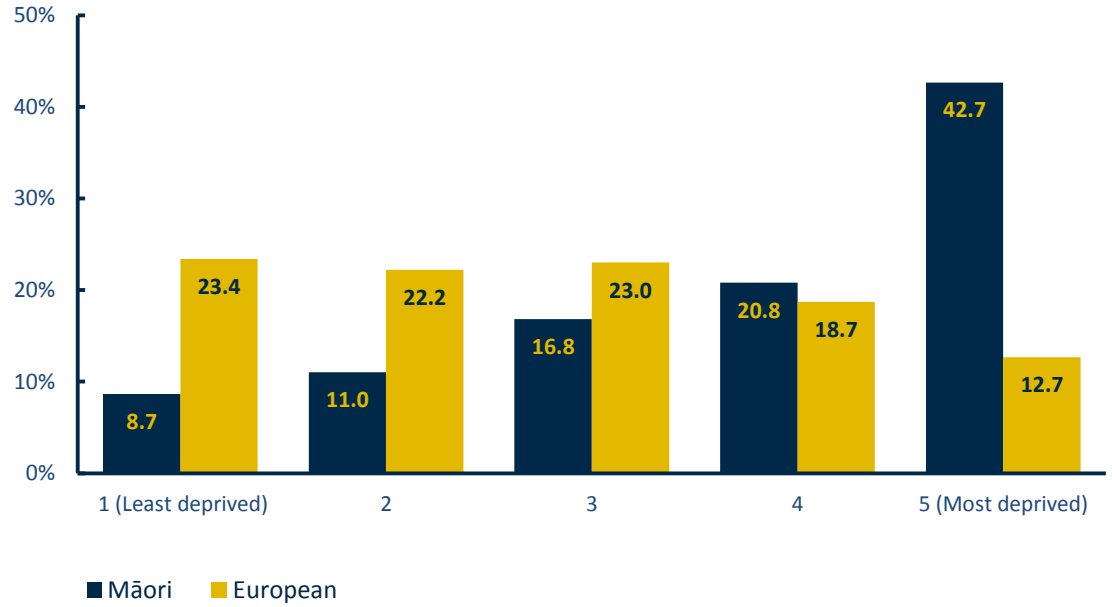


Figure 12.2: Māori and European population: profile of NZDep2013 quintiles (NZCASS 2014)



All things being equal

The effect of contributing factors can be removed by standardisation, which re-weights the factors of the Māori and European populations to give the same structure as the combined Māori and European population that is, we are trying to answer the question: What would the victimisation risk gap be between Māori and Europeans if they had a similar age and deprivation structure? This means that any remaining differences in victimisation between Māori and Europeans would then not be due to the age and deprivation differences between the 2 populations.

This approach to control for multiple factors at once is termed *multiple standardisation* – a term and analysis technique used, for example, in Australian Bureau of Statistics (2014). For this analysis, age and deprivation are the factors to standardise by, which means we are analysing the effect of what if both Māori and Europeans had the same age structure and deprivation. The process to do this is to 'weight up' or 'weight down' the responses to give the same profile across the combined population. For example, since there are more Māori aged 15–19 years, these responses would be 'weighted down' and the European 15 –19-year-old responses would be 'weighted up' so the age and deprivation profiles are the same. Then the victimisation rates are re- calculated on the re-weighted dataset.

For people that identified as both Māori and European, this analysis halved their original weight so they would only contribute once in the analysis (half as Māori and half as European). If this wasn't done, people who identify as both ethnicities would be analysed twice, which would in effect decrease the difference between Māori and Europeans as more weighted individuals would be alike.

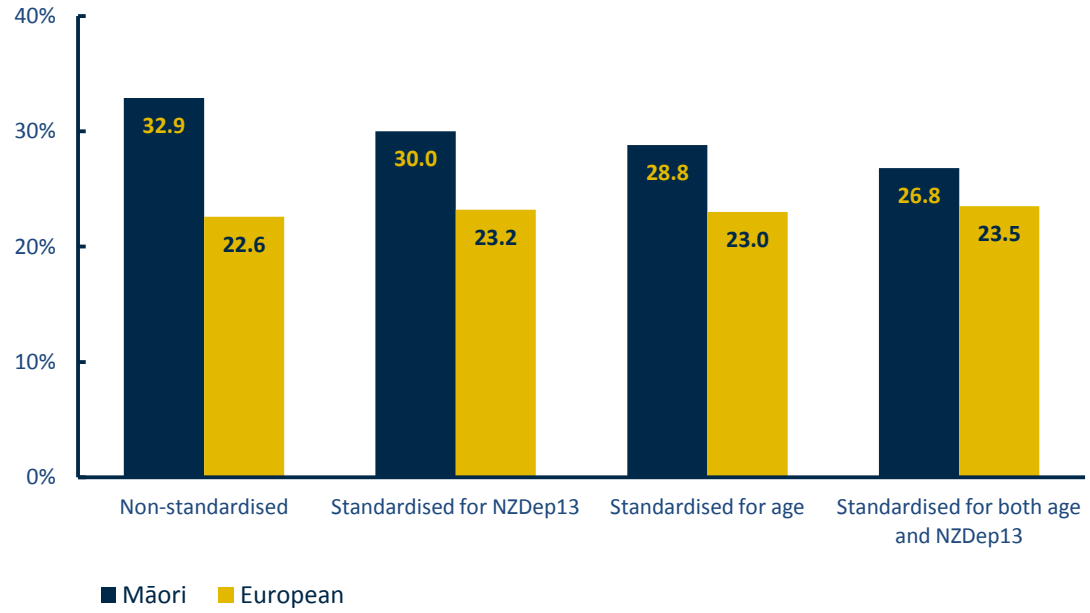
Multiple standardisation results

After adjusting for a combination of age and deprivation, the gap in victimisation risk between Māori and Europeans fell from 10.6 percentage points to 3.7 percentage points. This difference remains statistically significant. These results are summarised in Table 12.1 and Figure 12.3, along with the individual contribution to decreasing the victimisation risk gap for each of age and deprivation.

Table 12.1: Multiple standardisation results

	Māori	European	Difference (percentage points)
Not Standardised	32.9%	22.6%	10.3
Standardised for multiple factors			
Both age and NZDep13	26.8%	23.5%	3.3
Standardised for individual factors			
Age	28.8%	23.0%	5.8
NZDep13	30.0%	23.2%	6.8

Figure 12.3: Standardised victimisation rates for Māori and European



This analysis tells us that once controlling for age and deprivation, Māori are still more likely to be victimised than Europeans; however, over two-thirds of the victimisation risk difference can be accounted for by age and deprivation alone.

This analysis only accounts for 2 of the structural differences between the Māori and European populations. There will be other factors that may or not be measured in the NZCASS that will impact the victimisation risk differences between the 2 populations.

Logistic regression

The NZCASS data tables present a wide range of cross-tabulations of 2 or more data items. However, this presentation does not take into consideration the relationship between different factors. For example, young people are more likely to be students and live with flatmates, while people aged 65 years or more will be more likely to be retired. So if a factor comes through as statistically significant compared to the New Zealand average, such as being a student, it may not be that because someone is a student they are at higher risk of victimisation, but rather because student status is correlated with being younger. Due to this, it is difficult from the factor tables to assess which factors are directly related to victimisation, and which factors are secondary factors related to victimisation only through correlation to another factor. Regression was used as a modelling technique to deal with this overlap, where the model can control for holding other variables constant. The modelling process selects the most predictive combinations of many overlapping factors, and estimates their effect.

Models

The regression models were built for the probabilities of the following response variables.

Victimisation of:

- all offences
- interpersonal violence
- interpersonal violence by an intimate partner
- burglary
- thefts and damage offences

- vehicle offences.

As each of these response variables are whether someone was victimised or not, the models were fitted as logistic regression models. This is a widely used method for explaining binary response variables.

These models model whether someone was victimised or not, rather than the number of times they were victimised. This was done since whether someone was victimised or not is more stable, whereas the number of times someone was victimised can be affected by a few highly victimised people. But note – slightly different results might have emerged based on predictors for the number of times someone was victimised.

Explanatory variables

Table 12.2 summarises the explanatory variables included in the initial models.

Table 12.2: Explanatory variables included in initial regression models

Household offences (Burglary, vehicle offences)	Personal offences (All offences, interpersonal violence, intimate partner violence, theft and damage offences)
Personal factors	
	Sex
Age ◊	Age ◊
Ethnicity	Ethnicity
	Partnership status
Economic factors	
	Employment status
	Financial stress: limited to buy item for $300 ◊
Financial stress: can meet unexpected expense	Financial stress: can meet unexpected expense
	Personal income ◊
Household income ◊	Household income ◊
Household factors	
Household composition	Household composition
Tenure and landlord type	
Geographic factors	
Urbanisation	Urbanisation
Region	Region
Other factors	
Average rating of social disorder ◊	Average rating of social disorder ◊

◊ Included in model as a continuous variable. All other variables treated as categorical.

Note: For the burglary and vehicle offences model, it is somewhat artificial to analyse household offences against personal characteristics (such as age and ethnicity), since this depends on which respondent in the household was selected for the interview. For this reason characteristics such as sex and employment status were not included in the household models. Age, ethnicity and financial stress are also personal factors, but these are considered more homogenous amongst household members than other personal factors. However,

caution is advised when interpreting these personal factors in the household models, and the interpretation is that the characteristic reflects the average profile of household members.

These explanatory variables were selected from the standard range of demographic and geographic data items included in the NZCASS data tables. From this standard set, there are 3 exceptions:

1. NZDep2013 quintiles were excluded because they are derived from multiple measures of deprivation, some of which were also included in the model. It is preferable for un-derived variables to be retained as they are easier to interpret.
2. Legally registered relationship status was excluded as partnership status is derived from this.
3. A social disorder rating was included. This was calculated from the 6 aspects of neighbourhood crime problems. Respondents were asked to rate each of the 6 aspects from '1 – A very big problem' through to '4 – Not a problem at all'. The disorder rating was calculated for each respondent as the average of the codes for all 6 aspects. The scale was reversed so that a higher rating indicated greater disorder.

Model specification

From this list of explanatory variables the model that was fitted was:

$$victimised \sim personal + economic + household + geographic + other\ factors$$

where victimised is a binary response where 1 = victimised, 0 = not victimised.

The categorical variables in Table 12.2 were included in the model as dummy (indicator) variables. This means individual variables were included for each category within a factor, which takes the value of 0 or 1 to indicate the presence or absence of a category (eg has value 1 to represent 'employed' and has value 0 to represent everyone else). The alternative was to include the factor categories as 1 variable (eg employed, unemployed, etc), but dummy coding was done to assist interpretation of the victimisation odds ratio for that category compared to the rest of the population, rather than that category compared to a specified baseline.

When categorical variables are converted into multiple dummy variables, they can exhibit redundancy. For example, the dummy variable of 'Rest of South Island' can be identified as none of the other regions (Auckland, Wellington, Rest of North Island, Canterbury). This fails an assumption of logistic regression, and hence when this occurred, 1 less category than the number of categories was excluded from the initial model. The decision on which category to exclude was a conceptual one based on which category was most similar to the New Zealand average or most met the project tea m's research needs. For example, 'other multi-person household' was the household composition category excluded since it is a smaller category that wasn't significantly different from the New Zealand average. Similarly, 'Rest of South Island' was the region category excluded, since the NZCASS research needs were primarily to focus on the other regions. This is not to say these excluded categories are not important – these categories are still being represented in the model except more indirectly as they are represented as the '0' group of every dummy within that factor.

A number of variables were treated as continuous in the model (denoted with a ◊ in Table 12.2) in order to keep the natural ordering of these variables in the model. The categories used for these variables varied from that presented in the data tables to ensure that similar increments were used. For example, age is presented in the data tables as a mix of 5-year and 10-year age groups, and with the upper group of 65 years and over. However, for the regression, age was included in evenly sized 5-year age groups, and the upper group was 75 years and over.

Interaction terms were included for selected variables (including age/ethnicity, employment status/ethnicity and household composition/financial stress); however, it was not appropriate to fit all possible interactions due to the number of explanatory variables and sample size restrictions. Quadratic terms of personal and household income were included for selected offence models where there was assessed to be a non-linear relationship with victimisation (eg when lower income has higher rates of victimisation, middle income has lower rates of victimisation, and then higher

income has higher rates of victimisation). The assumption of linearity was valid for the remaining continuous variables.

The regression models were fitted unweighted. All other NZCASS analysis was conducted using weights to compensate for imbalances in the survey profile relative to the target population. However, with this regression analysis, the intention differs in regard to prediction. With other NZCASS analysis we aim to describe victimisation rates in 2013 (ie the 'how many'), whereas this regression analysis aims to provide an understanding of the predictors of victimisation (ie the 'who'). Hence it was decided an unweighted model was more appropriate, and this is consistent with the modelling approach in previous NZCASS iterations.

Furthermore, there were practical considerations in that SAS does not implement a step-wise backwards weighted regression using proc survey logistic. There were other, less important, reasons for preferring unweighted estimates such as robustness with respect to extreme weights. The differences between the weighted and unweighted results are briefly discussed under 'Weighted regression'.

Missing values in the explanatory variables (such as refusal or 'don't know' responses) were imputed to prevent the entire record being dropped during the modelling process. The following imputation method was used:

1. First, missing records were matched with potential donors using known characteristics of age, deprivation and ethnicity.
2. The donor's value of the missing cell were then assigned to the respondent. In cases where there were multiple donors (most of the time), 1 potential donor was randomly selected from the pool. In the very small number of cases where there were no donor matches on age, deprivation and ethnicity, the donors were matched on age and deprivation alone (every respondent as matched just on the 2 characteristics).

Influential observations were assessed, and a small number of observations (up to 6) were removed from each model.

Model selection

The regression analysis was a step-wise backwards elimination method, which starts with the full model and drops 1 variable at a time until all remaining variables are statistically significant. This is a widely used variable selection technique, but does have the known limitation that constructed confidence intervals are too narrow. This approach was selected since there are a large number of candidate explanatory variables, coded into 35 main effects plus interactions for the personal offences models. This works out to be billions of different model combinations, even without considering interactions. Hence the step-wise elimination method is a relatively quick method that can be automated to evaluate a range of explanatory variables.

As the NZCASS dataset has 100 imputations, the regression models were built on the 100 imputation datasets separately. The final model was specified by combining the results from the 100 individual models using the standard Rubin combining rules (see 'Analysis Methods' for further information). Only explanatory variables that were retained in at least 40 of the 100 models were included in the final model. This threshold of 40 was used because from an analysis of frequencies, this was a natural separation point where there was a cluster of variables only included in a few or dozen models, while the remaining were included in most or all of the models.

The significance level was set at 95%. Other significance levels were considered, such as 90%, but the 95% level was used for 2 main reasons:

1. 2014 NZCASS reporting was done at the 95% level
2. the modelling aimed to determine the 'best' predictors of victimisation (ie not including too many variables that may be included with a lower threshold).

While considering the need to not exclude potentially important predictors, this second consideration was balanced with the decision to include the explanatory variables retained in 40 of the 100 imputation final models. The threshold was slightly lower with the aim to be conservative, and to ensure the final pooled model had a similar number of explanatory variables to the final individual models.

Interpreting results
The logistic regression model expresses the logarithm of the odds ratio of being a victim as a linear combination of the explanatory variables. However, for ease of interpretation, the odds ratios have been presented by taking the exponential of the log odds ratios.

The odds ratio represents higher or lower odds of victimisation, while controlling for the other factors. An odds ratio greater than 1 indicates higher odds of victimisation, where a number less than 1 indicates lower odds of victimisation when compared to the reference group.

Odds are not the same as probabilities – both are numerical measures of how likely an event is to occur, but have different interpretation.

Probabilities
Probabilities are the chance, or risk, that something will occur. For example, if we use 2013 statistics to estimate future risk, there is a 9.5% probability that a household in Auckland will be burgled in the next year. Similarly, there is a 7.4% probability that a household outside Auckland will be burgled.

Odds
Odds can be calculated from probabilities where odds = probability/(1 − probability). For example, the odds of a household experiencing a burglary in Auckland in the next year are 0.105 (0.095 /1 − 0.095). Similarly, the odds of a household not in Auckland experiencing a burglary in the next year are 0.080 (0.074 /1 − 0.074).

Odds ratio
The odds ratio is the odds relative to another group. For example, the odds ratio of a household experiencing a burglary in Auckland compared to the rest of the country is 1.31 (0.105/0.080).

Interpretation
This means the odds of experiencing a burglary in Auckland is 31% higher than the odds for experiencing a burglary in the rest of the country.

The interpretation of odds ratios for categorical and continuous explanatory variables differs. For continuous variables, the odds ratio represents the victimisation odds change with 1 unit increase in that characteristic holding other variables constant. For example, the unit increase in age is in 5-year bands. Table 3 shows the odds ratio for age is 0.9 for violent interpersonal offences. This means the odds of victimisation are 10% lower for every 5-year age increase, holding other factors constant.

For categorical variables the odds ratio represents the victimisation odds compared to the reference group, holding the other variables constant. The reference group is as follows:

- For cases where only 1 category of that variable is retained in the final model or for the multiple response category of ethnicity, the reference group is the rest of the population. For example, Table 12.3 shows that ‘1 parent with children’ households have an odds ratio of 1.31 for violent interpersonal offences. This means the odds of being a victim of violent interpersonal offences are 31% higher than that of the rest of the population when other factors are held constant.
- For cases where there are 2 or more categories of that variable retained in the final model, the reference group is everything outside those categories. For example, Table 12.3 shows the odds ratio for the ‘all offences’ model for Auckland is 1.18, and that Auckland and Canterbury were the 2 regions retained in the final model. This is therefore interpreted as the odds of being victimised in Auckland are 18% higher than compared to odds of someone not in Auckland nor Canterbury being victimised.

Final model results

Table 12.3 summarises the odds ratio and the corresponding CIs for the best (final) models.

Table 12.3: Final logistic regression model results

Model	All offences	Burglary	Thefts and damage offences	Vehicle offences	Violent interpersonal offences	Intimate partner violence
Odds ratio [95% CIs]						
Intercept	0.09 [0.06 – 0.13]	0.02 [0.01 – 0.03]	0.02 [0.00 – 0.05]	0.03 [0.02 – 0.06]	0.07 [0.04 – 0.12]	0.11 [0.05 – 0.27]
Categorical variables						
Sex: Female						1.36 [1.05 – 1.75]
Ethnicity: European			1.31 [1.03 – 1.66]		1.28 [1.06 – 1.54]	
Ethnicity: Māori	1.24 [1.09 – 1.42]	1.41 [1.17 – 1.69]	1.64 [1.04 – 2.60]		1.56 [1.30 – 1.88]	1.56 [1.15 – 2.12]
Partnership status: Partnered – legally registered					0.67 [0.55 – 0.81]	0.51 [0.36 – 0.72]
Partnership status: Partnered – not legally registered	1.20 [1.04 – 1.40]				1.40 [1.14 – 1.72]	1.64 [1.20 – 2.24]
Employment status: Retired	0.74 [0.58 – 0.94]					
Financial stress: Can meet unexpected expense			0.90 (Ł) [0.68 – 1.19]		0.73 [0.59 – 0.90]	0.57 [0.42 – 0.77]

Model	All offences	Burglary	Thefts and damage offences	Vehicle offences	Violent interpersonal offences	Intimate partner violence
Financial stress: Can meet unexpected expense * Household comp: 1 parent with child(ren) and other person(s)			3.33 [1.13 – 9.81]			
Household comp: 1 parent with child(ren)	1.31 [1.09 – 1.58]	1.40 [1.06 – 1.85]			1.32 [1.03 – 1.68]	
Household comp: 1 parent with child(ren) and other person(s)			0.48 (Ł) [0.19 – 1.23]			
Household comp: Couple only					0.70 [0.55 – 0.89]	
Household comp: Couple with child(ren)		1.27 [1.04 – 1.56]			0.73 [0.59 – 0.90]	
Household comp: Couple with child(ren) and other person(s)		1.65 [1.07 – 2.55]				
Tenure and landlord type: Rented – government (local and central		1.41 [1.05 – 1.90]				
Urbanisation: Main urban area	1.25 [1.07 – 1.46]	1.41 [1.15 – 1.74]		1.34 [1.04 – 1.72]		
Urbanisation: Secondary urban area	1.41 [1.09 – 1.83]		1.79 [1.14 – 2.81]		1.41 [1.05 – 1.91]	
Region: Auckland	1.18 [1.02 – 1.35]			1.67 [1.32 – 2.11]		
Region: Rest of North Island			0.70 [0.55 – 0.90]			
Region: Canterbury	1.23 [1.02 – 1.49]					

Model	All offences	Burglary	Thefts and damage offences	Vehicle offences	Violent interpersonal offences	Intimate partner violence
Continuous variables						
Age (Increasing age)	0.93 [0.91 – 0.96]	0.96 [0.93 – 0.99]	0.96 (Ł) [0.92 – 1.01]	0.91 [0.87 – 0.94]	0.90 [0.87 – 0.92]	0.88 [0.83 – 0.93]
Age* Ethnicity: Māori (Increasing age)			0.93 [0.87 – 0.99]			
Financial stress: Limited to buy item for $300 (Increasingly limited)	1.08 [1.04 – 1.13]		1.12 [1.03 – 1.21]		1.09 [1.03 – 1.16]	
Personal income (Increasing income)	1.06 [1.01 – 1.12]					0.87 [0.79 – 0.97]
Personal income squared (See note below)			1.02 [1.00 – 1.03]			
Average rating of social disorder (Increasing disorder)	1.86 [1.69 – 2.04]	1.98 [1.74 – 2.25]	2.20 [1.89 – 2.57]	1.81 [1.54 – 2.12]	1.71 [1.51 – 1.93]	1.50 [1.25 – 1.80]

Notes:

Only variables that were retained in the final models are presented.
Ł Not significant but included in final model as interaction term included. Interaction term – that is, the simultaneous combination of the 2 variables.

Decreasing victimisation odds (but note the interpretation of interaction terms).

Increasing victimisation odds (but note the interpretation of interaction terms).

The interpretation of interaction terms needs to be done alongside the main effect terms. For example, for theft and damage offences the interaction age*Māori is negative (odds ratio 0.93) and the age main effect is negative (odds ratio 0.96), whereas the Māori main effect is positive (odds ratio 1.64). When calculating the combined probability for Māori, the change in the negative contribution is greater than the change in the positive contribution. Hence for Māori, the effect of increasing age is still lower victimisation odds.

The interpretation of quadratic terms (in this case personal income squared) is that since the term is greater than 1, the relationship with victimisation is 'U' shaped, in that people with lower and higher personal income have higher odds of victimisation, whereas people with middle income have lower odds of victimisation.

Note that the reverse interpretation can be applied. For example, people in a not-legally registered partnership status have higher odds of victimisation for all offences than the rest of the population. Conversely, this can be interpreted that people who are in this 'res t of the population' group (ie legally registered partnerships and non-partnered people) have lower odds of victimisation. This is particularly important to consider when some categories were excluded due to a voidance of dummy variable redundancy (see Model Specification).

Model fit

The predictive power of the logistic regression models has been measured by a statistic called the area under the Receiver Operating Characteristic (ROC) curve. If the model is weak at distinguishing victims from non-victims, the ROC statistic will have a value around 0.5 (no better than a coin toss). Whereas, if the model is perfect at distinguishing victims from non-victims, the statistic will have a value of 1.0. Table 12.4 shows the ROC statistics for each model.

Table 12.4: Model fit statistics for each regression

Model	ROC statistic
All offences	0.685
Burglary	0.686
Thefts/damage offences	0.699
Vehicle offences	0.687
Violent interpersonal offences	0.726
Intimate partner violence	0.756

As most of the ROC statistics range between high 0.6 to mid 0.7, this shows the models are helpful but do not have perfect explanatory power. There still remain other unmeasured factors or perhaps random behaviour that puts some people/households more at risk of victimisation than others. The household offences models have lower predictive fit than the personal offences, indicating there are more unmeasured factors (such as presence of alarms/CCTV) or random behaviour for these offences.

Caveats

There are a number of caveats to be aware of when using regression results.

- **Correlation does not prove causation:** Not all possible drivers of victimisation are included in the regression models, since the NZCASS does not collect all possible characteristics or predictors of victimisation. Variables may be retained in the final model only because they are related to an unmeasured variable. For example, age was retained in all the final models, but it may not be someone's actual age that puts them at risk, but rather the way they socialise, how they live and the places they go. For this reason, the statistical models do not provide a perfect explanation of what predicts victimisation.
- **Collinearity:** Several variables may be correlated, but just 1 may be retained in the final model that is most closely related to victimisation. This does not mean that the other variables have no importance at all, but rather, it is not the 'best' predictor of victimisation.
- **Sample size:** An explanatory variable may not be retained in the final model simply due to not having the sample size to support its inclusion. This is not to say that it is not important in predicting victimisation. For example, this may affect the findings for Pacific peoples since the sample size is smaller than that for Māori or European ethnicity.

Weighted regression

Unweighted logistic regression models were used for the regression analyses presented in this report. The same models were re-run (using proc surveylogistic) with weights, stratification variables of region and urbanisation, and cluster variable of meshblock specified. The results of the unweighted and weighted models were compared.

Most importantly from this comparison, the differences between the odds ratios from the unweighted and weighted models were minor. The estimates were very similar for most explanatory variables, or where they did differ, the direction (ie higher or lower odds ratio) was consistent. Only a small number of estimates differed largely in terms of magnitude, and when that occurred it could

be explained by the variable not being statistically significant in either the weighted or the unweighted model. Hence it can be concluded that overall, weighting made immaterial difference to the odds ratios.

While the effect on odds ratios was minor, the use of weights would have changed the model variable *selection* results. Table 12.5 summarises which explanatory variables that were retained in the final unweighted model did not meet the 95% level criteria (had p values greater than 0.05) when run weighted.

Table 12.5: Explanatory variables that do not meet the 95% threshold with a weighted model

Model	Explanatory variables
All offences	Personal income Household comp: 1 parent with child(ren) Partnership status: Partnered – not legally registered Urbanisation: Secondary urban area
Burglary	Household comp: Couple with child(ren) Household comp: Couple with child(ren) and other person(s) Tenure and landlord type: Rented – government (local and central)
Thefts/damage offences	Ethnicity: European Financial stress: Can meet unexpected expense Household comp: 1 parent with child(ren) and other person(s) Region: Rest of North Island Urbanisation: Secondary urban area
Vehicle offences	Region: Auckland
Violent interpersonal offences	Ethnicity: European Partnership status: Partnered – legally registered Partnership status: Partnered – not legally registered Financial stress: Can meet unexpected expense Financial stress: Limited to buy item for $300 Household comp: Couple with child(ren) Household comp: 1 parent with child(ren) Urbanisation: Secondary urban area
Intimate partner violence	Financial stress: Can meet unexpected expense Personal income Partnership status: Partnered – not legally registered Sex: Female

Note: This table does not show which alternative variables (if any) would have been retained in a weighted model using the same step-wise backwards elimination method.

Most of these were close to the threshold of 95% (p values of 0.05), but some were larger indicating there is sample design effect and clustering of those variables. Particularly for violent interpersonal offences, the potential variable identification is quite different for the weighted and unweighted models. This could be explained by a small number of respondents with large weights having a large effect on the retention of variables, and/or if violent interpersonal offences are highly concentrated, then considering strata and clustering would effectively reduce the sample size, and hence reduce model power to determine predictors of victimisation. But in general, this is provided for user information and for the reasons discussed under 'Model specification', the unweighted model has been assessed as the more preferred model to be used to provide an understanding of predictors of victimisation.

The Gini coefficient

This part of the advanced statistical methods section uses a 'Gini coefficient' to provide a measure of the distribution of victimisation. It is appealing as a statistic since it summarises the distribution in a single summary statistic, enabling comparisons over time and between groups to be captured succinctly. This can be used to see whether victimisation is becoming more or less equal over time, and also to assess whether the distribution of some offences is more unequal than others.

The Gini coefficient is a summary measure of inequality, used widely in analysing the distribution of income or wealth. Extending this to victimisation, it is a measure where:

0 = perfect equality – where all members of the population experience the same amount of victimisation

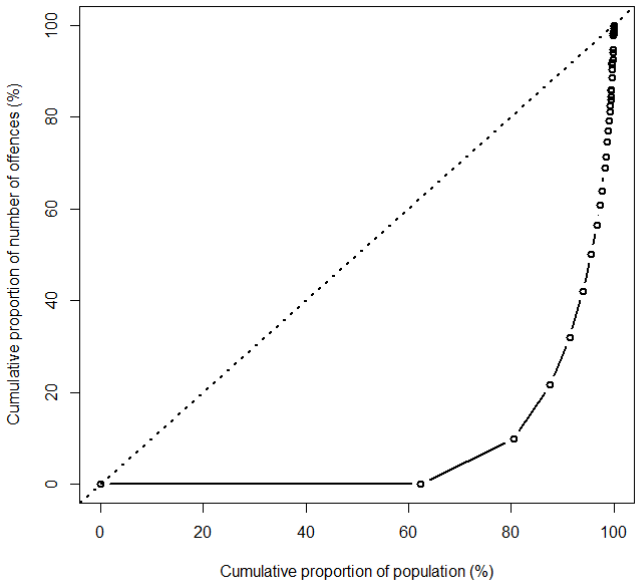
1 = total inequality – where only 1 person/household experiences all the victimisations, and everyone else experienced none

The smaller a Gini coefficient, the more even the distribution of victimisation, and the larger a coefficient, the more concentrated victimisation is amongst a group of highly victimised people.

Mathematics of the Gini coefficient

The Gini coefficient is calculated in reference to the Lorenz curve – a graph with the horizontal axis showing the cumulative proportion of the population ranked according to the number of offences experienced, and the vertical axis showing the corresponding cumulative proportion of the number of offences they have experienced, as shown in Figure 12.4.

Figure 12.4: Example Lorenz curve



If victimisation was evenly distributed amongst the population, then the Lorenz curve would be the diagonal (line of equality). The value of the Gini coefficient is the ratio of the area between the Lorenz curve and the diagonal to the area of the entire diagonal triangle.

Calculations for the NZCASS

This area ratio is equivalent to calculating the Gini coefficient through a direct formula. This was implemented in the NZCASS analysis as follows:

1. Calculate weighted frequencies and percentage of number of offences experienced for all number of offences i.

2. Calculate the cumulative number of offences experienced, and the cumulative percentage of population (x).
3. Calculate the cumulative percentage of offences experienced (y).
4. The Gini is then calculated as:

$$\sum[x_i \times y_{i+1}] - \sum[x_i \times y_{i+1}]$$

The final Gini is the last row (ie where y = 100).

Populations

The populations used for the Gini calculations varied for the comparisons across time and comparisons between offence groups:

- *Comparisons across time*: The population is the total adult population (ie those who experienced no offences were included in the calculations)
- *Comparisons between offence groups*: The population is the victim population (ie those who experienced no offences were excluded in the calculations).

For the comparisons across time, we want to see how the distribution of victimisation has changed for all adults, so the large percentage of adults who did not experience any victimisation are an important part of this story. However, for comparisons between offence groups, we are more interested in the distribution of victimisation for victims (i e excluding the non-victims), and since a large portion of the population would have experienced none of that offence group, it would skew the Gini statistic upwards.

Gini results

The Gini results are presented in the NZCASS Main Findings Report, and also included here along with the Lorenz curve graphs to assist interpretation.

Table 12.6: Gini results – across time

	2006	2009	2014
Gini coefficient	0.810	0.814	0.886‡

‡ Statistically significant difference across time at the 95% confidence level from both 2005 to 2013 and 2008 to 2013.

The following Lorenz curve plot (Figure 12.5) assists interpretation of why the Gini has increased in 2014 – that is, because there are now more people who have not been a victim of crime, but there is a small group of people who remain highly victimised, then the distribution of victimisation is now more unequal than in previous years.

Figure 12.5: Lorenz curves – all offences by year (for adults)

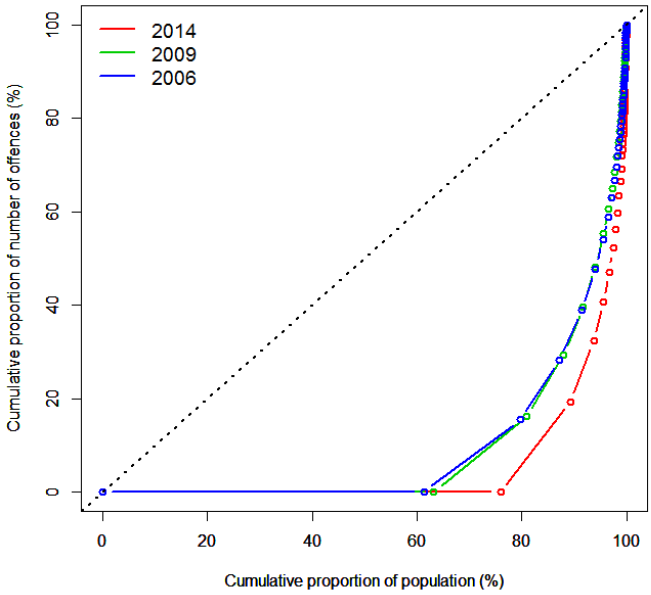


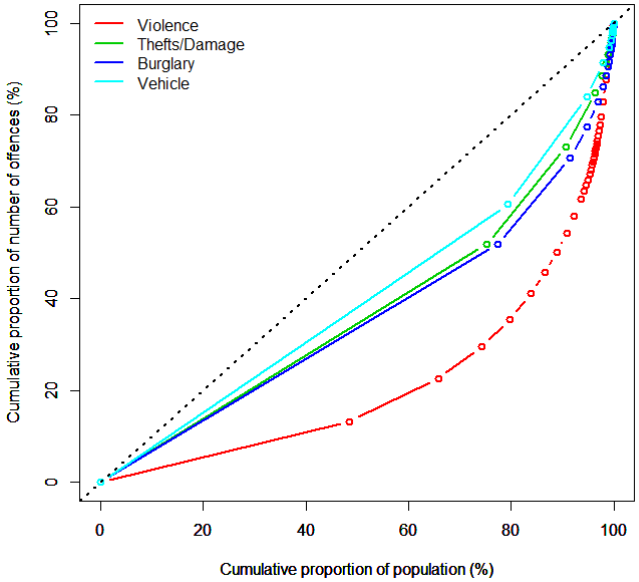
Table 12.7 and Figure 12.6 present the Gini results by offence group for 2014.

Table 12.7: Gini results – between offence groups – 2014

	Gini coefficient
Interpersonal violence	0.566
Burglary	0.289*
Thefts and damage	0.262*
Vehicle offences	0.203*

* Statistically significant difference to the interpersonal violence Gini coefficient at the 95% confidence level.

Figure 12.6: Lorenz curves – 2014 by offence (for victims)



Comparing NZCASS and official Police statistics

Introduction to Police statistics

Traditionally Police have counted offences ('historic Police offence statistics'); however, this series of statistics ended on 1 April 2015. The historic Police offence statistics have now been replaced by the Police 'recorded crime victim statistics' (RCVS).

Comparisons between NZCASS and Police statistics have been made based on the historic Police offence statistics (rather than RCVS) for 2 main reasons:

1. The recall period for the 2014 NZCASS was 1 January 2013 to the date of interview (February – June 2014). The RCVS series only started from July 2014, meaning the time periods were incomparable.
2. To look at trends with previous surveys, comparisons with Police needed to be made based on the same statistics.

In the next iteration of the NZCASS, comparisons will be made against the Police-recorded crime victim statistics, which are likely to be more comparable with NZCASS estimates.

Comparing NZCASS and historic Police offence statistics

The NZCASS is a sample survey whereas the data used in historic Police offence statistics was collected through administrative processes (information on offences was collected and stored electronically in the Police National Intelligence Application (NIA)).

Table 12.8: Method and content differences between Police and NZCASS statistics

	Police statistics	NZCASS statistics
Method	Census count of all offences recorded in the Police NIA.	Sample survey of 6943 New Zealand residents (aged 15 years and over).
Content	Historic Police offence statistics count offences recorded by Police. Counts are based on the number of offences reported by the public or detected by Police.	Asks people about their feelings of safety and experiences of crime (along with a range of other questions).

Comparing these different types of data will affect the completeness and precision of the statistics produced.

Table 12.9: Types of error in Police and NZCASS statistics

	Police statistics	NZCASS statistics
Sampling error	Administrative data is not subject to sampling error.	Because not all New Zealanders are interviewed about their experiences of crime, NZCASS statistics are subject to sampling error.
Respondent reporting	Administrative data holds actual records of incidents that are reported to or discovered by Police, and recorded in Police systems.	The NZCASS asks if Police came to know about the incident but there is currently no accurate and robust way of verifying this with Police records. This means a respondent may say that the incident was reported to Police but it may not have been.

Classifying offences

As a general principle, offences in the NZCASS are coded:

- in line with current legal theory
- in line with current Police recording procedures.

Despite this, an exact match is unlikely because Police may make different judgements when deciding:

- whether to record an incident as an offence
- which category it should be placed in.

This means that some offences may have been placed into a different offence category, which will affect comparability.

Same types of offences covered

The NZCASS does not collect information on all types of offences, whereas Police statistics do. For example, the NZCASS does not cover 'victimless crimes' (such as drug offences) and offences where the primary victim cannot participate in the survey (such as manslaughter or homicide), while Police statistics exclude some offences that the NZCASS includes, such as offences not reported to Police.

Do statistics include the same things?

Police statistics include a range of things that NZCASS statistics do not. The most common differences that affect comparability are age and commercial targets.

Table 12.10: Age and commercial target differences between Police and NZCASS statistics

	Police statistics	NZCASS statistics
Age	Includes victimisations of those aged under 15 years.	Excludes victimisations of those aged under 15 years.
Commercial targets	Includes crimes committed against commercial targets, such as business vehicles and commercial premises.	Excludes crimes against commercial targets. (Note that commercial crime may sometimes be included because descriptions of events may be hard to understand.)

Police recorded crime victim statistics (RCVS)

Police victim statistics can include a range of crimes that NZCASS statistics do not.

Table 12.11: Offence differences between Police and NZCASS statistics

	Police victim statistics	NZCASS statistics
Death-related offences	Includes: • murder • manslaughter • driving causing death.	Excludes offences where the victim is dead.
Blackmail and extortion	Includes blackmail and extortion.	While this can be collected as part of NZCASS, blackmail and extortion are not counted.
Theft from retail premises	Includes crimes where a business is the victim.	Excludes crimes against commercial targets.

Table 12.12: Relationship differences between Police and NZCASS statistics

	Police victim statistics	NZCASS statistics
Ex-partners and ex-boyfriends/girlfriends	Included under 'Non-family member'.	Included with 'Intimate partner' categories and 'Family' categories.
Children	Included within statistics.	Excluded from the NZCASS as those under 15 years old are not part of the survey.

Counting and recording offences

Police may make different judgements when deciding whether to record an incident as an offence. Table 12.13 provides a range of examples where this might happen and how NZCASS statistics differ.

Table 12.13: Counting rule differences between Police and NZCASS statistics

	Historic Police offence statistics	NZCASS statistics
Series of the same type of offence	In some cases, not all incidents in a series are counted. For example, in cases of partner assault, when offences are repeated by the same offender against the same victim over a period of time, Police may count this as a single offence depending on the circumstances (such as if they don't have specific details on individual offences).	The NZCASS counts each incident. The NZCASS estimates for offences like assault will be higher than the Police statistics.
Series of different offences	Where a series of offences have been committed, not all of these may be counted: • where 1 offence is part of another • 1 offence is the means to committing another • where the offences are relatively • minor • where there is insufficient • information (who, time, place, type what happened) to clearly distinguish between offences.	The NZCASS counts multiple offences in the same way as Police do wherever possible.
Dealt with as a crime	Not all incidents reported to Police end up recorded as offences. An offence is not recorded if, at any stage after the initial report, Police judge that an offence (in law) probably did not occur. In addition, where offences are minor and Police and the victim believe attendance is sufficient to resolve the situation, these offences may also not be recorded (eg the case of some threats).	The NZCASS counts each offence based on whether it meets the legal definition of a given crime based on the information provided by victims. Statistics are not influenced by whether a victim wants an incident to be treated as an offence or not. As such, NZCASS estimates are likely to be higher in some cases.
Enough evidence	Police may not record offences reported to them when: • there is not enough evidence to establish whether an offence happened or not • there is evidence that the offence didn't happen • they do not consider the person reporting the offence to be credible.	Within the NZCASS: • where there is not enough • information to establish whether an offence occurred or not, the incident is coded as out of scope • the NZCASS only gathers information from the victims' perspective, so no alternative views can influence count estimates • no judgement about the credibility of the victim is made.

Categories used to report the statistics

Survey information is sometimes grouped differently from administrative data.

Table 12.14: Grouping of data differences between Police and NZCASS statistics

	Description
Sample constraints	Survey data is sometimes grouped to make sure that statistics are based on large-enough sample sizes and are robust. Administrative data is often not subject to the same restrictions, given the amount of data available. Because of this, groupings used to report survey statistics will not always align directly with the higher-level Australian and New Zealand Standard Offence Classification (ANZSOC) because the sample is not big enough.
Data availability	In some cases, not all elements of a given ANZSOC classification are available in the NZCASS. For example, the level 1 ANZSOC classification for 'sexual assault and related offences' includes elements of sexual violence that are out of scope for the NZCASS, such as information about 'child pornography' and 'non-assaultive sexual offences against a child'. Likewise, for some offence categories, the Police figures include a large number of different offences that would be difficult to untangle to match the coverage of the NZCASS.
User needs	NZCASS analysis and reporting needs to be useful and relevant to a wide range of users. Some ANZSOC classifications may not be the best way to report information for the majority of these users.
Consistent reporting	In some cases, the groupings used to report survey statistics need to be consistent with previous iterations of the research. This is done because some groupings are used to monitor specific trends. These groupings may not align directly with ANZSOC classifications.

Offences that can be compared with official Police statistics

Because of the complications discussed above, there is only a subset of offences that can be compared with Police statistics. This subset includes:

- thefts of vehicles
- thefts from vehicles/vehicle interference
- burglary
- robbery/theft from the person
- assaults.

Note: A concordance between these NZCASS offences and ANZSOC groupings can be found in Appendix G.

Adjustments made

To make it easier to compare NZCASS and Police statistics, the following adjustments were made in 2014.

Table 12.15: Adjustments made to Police data to improve comparability with NZCASS

Offence	Adjustment made to Police data	Impact
Thefts of vehicles	Remove commercial vehicles.	Police recorded 18,016 thefts of vehicles in 2013. This was reduced by 4.5% to account for commercial vehicles as targets – down to 17,205.
Thefts from vehicles/vehicle interference	Remove commercial vehicles.	Police recorded 36,033 thefts from vehicles and vehicle interference in 2013. This was reduced by 3.9% to account for commercial vehicles as targets – down to 34,628.
Dwelling burglary	No adjustment needed. See Appendix G.	
Robbery/theft from the person	Remove victimisations against people 13 years and under.	Police recorded 4186 victimisations in 2013. This was adjusted to 3829.
Assaults	Remove victimisations against people 13 years and under.	Police recorded 37,291 victimisations in 2013. This was adjusted to 32,264.

How adjustments were made

The Police data was adjusted to account that the NZCASS only covers crime against those aged 15 years and over, and only includes private vehicles (discussed in Table 12.10).

The unadjusted Police data was obtained from the 'historic Police offence statistics' from NZ.Stat on the Statistics NZ website using the concordance included in Appendix G: NZCASS-ANZSOC concordance.

To obtain the data for the adjustments, a data request was placed with the Police Statistical Services Unit for 2013 calendar year using the provisional victimisation data. While the RCVS was only officially published from July 2014, the completeness of victim information in 2013 was estimated to be around 90%, hence seen as a source (albeit provisional) to improve on the adjustments made in previous iterations of the NZCASS. Note that the data requested differs with the published official statistics on the RCVS due to different counting rules. The data requested counts the number of victimisations for that offence category within the calendar year (to align to NZCASS counting rules). The RCVS in comparison counts a person/organisation for each criminal incident once per month for each ANZSOC division in which they are recorded as being a victim of an offence.

Table 12.16 shows what data was requested and how this data was used to make the adjustments:

Table 12.16: How adjustments were made to Police data to improve comparability with NZCASS

Adjustment	Data requested from the Police Statistical Services Unit (using the RCVS 2013)	How adjustment data used
Remove commercial vehicles	Whether victim is a person or an organisation, for 'theft of vehicles' and 'thefts from vehicles and vehicle interference'.	In the NZCASS, vehicle offences are treated as a household offence (1 household victimised) rather than counting the number of people within that household that were the victims of the vehicle offences. As such, the counting unit of the 'historic Police offence statistics' are more comparable to NZCASS. The victim level data obtained was used to pro-rata the offence level data, on the assumption that there was a similar percentage of vehicles owned by an organisation at the offence level and victim level. These adjustments were 4.5% for 'thefts of vehicles' and 3.9% for 'thefts from vehicles/vehicle interference'.
Remove victimisations against people aged 13 years and under	The age of victims for 'robbery/thefts from the person' and 'assaults' for the age groups '13 years and under' and '14 years and over'. **Note** that while the NZCASS includes only people aged 15 years and over at the time of interview, the victim recalls incidents that happened the year prior. Hence they could be aged 14 years at the time of incident.	In the NZCASS, robbery, thefts from the person and assaults are treated as a personal offence (1 person victimised). Hence for these offences, the counting unit of the RCVS is more comparable than the 'historic Police offence statistics'. The victim level data obtained by age groups was directly used as the adjusted version of the offence level data. In this case, the RCVS identified that there were 3829 victims aged 14 years and over of 'robbery/theft from the person' in 2013, and similarly 32,264 victims of 'assault'.

References

Atkinson, J, Salmond, C & Crampton, P. 2014. *NZDep2013 Index of Deprivation*. Wellington: Department of Public Health, University of Otago. http://www.otago.ac.nz/wellington/otago069936.pdf

Australian Bureau of Statistics. 2014. *Australian social trends, 2014*. http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/4102.0main+features72014

Australian Institute of Health and Welfare. 2011. *Principles on the use of direct age-standardisation in administrative data collections: For measuring the gap between Indigenous and non- Indigenous Australians*. http://www.aihw.gov.au/WorkArea/DownloadAsset.aspx?id=10737420130

Cochran, W G. 1977. *Sampling techniques*. New York: Wiley.

Gelman, A, Jakulin, A, Grazia Pittau, M & Su, Y-S. 2009. A weakly informative default prior distribution for logistic and other regression models. *Annals of Applied Statistics* 2(4): 1360–1383. http://www.stat.columbia.edu/~gelman/research/published/priors11.pdf

Graham, J W, Olchowski, A E & Gilreath, T D. 2007. How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science* 8(3): 206– 213. http://link.springer.com/article/10.1007%2Fs11121-007-0070-9

Kott, P S. 1998. *Using the delete-a-group jackknife variance estimator in NASS surveys*. NASS Research Report 98-01 (revised 2001). http://www.nass.usda.gov/Education_and_Outreach/Reports,_Presentations_and_Conferences/allreports/Using_the_Delete-a-Group_Jackknife_Estimator_in_NASS_Surveys.pdf

Lumley, T. 2004. Analysis of complex survey samples. *Journal of Statistical Software* 9(1): 1–19. http://www.jstatsoft.org/v09/i08/paper

Lumley, T. 2013. *Survey: Analysis of complex survey samples*. R package version 3.29-5. http://r-survey.r-forge.r-project.org/survey/index.html

Meng, X L. 1994. Multiple-imputation inferences with uncongenial sources of input. *Statistical Science* 9(4): 538–558. http://projecteuclid.org/euclid.ss/1177010269

O'Reilly, J. 2009. *Paradata and Blaise: A review of recent applications and research*. http://ibuc2009.blaiseusers.org/papers/7d.pdf

Reilly, J & Sullivan, C. 2008. Technical *report: New Zealand Crime & Safety Survey 2006*. Wellington: Ministry of Justice. http://www.justice.govt.nz/publications/global-publications/n/new- zealand-crime-and-safety-survey-2006/technical-report

Rubin, D B. 1987. *Multiple imputation for non-response in surveys*. New York: Wiley.

Salmond, C & Crampton, P. 2012. The development of New Zealand's deprivation index (NZDep) and its uptake as a national policy tool. *Canadian Journal of Public Health 103* (Suppl. 2): S7–S11. http://journal.cpha.ca/index.php/cjph/article/view/3195/2664

Seastrom, M, Kaufman, S & Lee, R. 2002. Explicit methods of imputing for item non-response. Appendix B3 in *NCES Statistical Standards, September 2002*. Washington, D.C.: National Center for Education Statistics, U.S. Department of Education. http://nces.ed.gov/statprog/2002/appendixb3.asp

Silverman, B W. 1986. *Density estimation for statistics and data analysis*. London: Chapman and Hall.

Statistics NZ. 1998. *Exploring the gap*. http://www.stats.govt.nz/browse_for_stats/income-and-work/Income/exploring-the-gap.aspx

Statistics NZ. 2013. 2013 Census *QuickStats about national highlights*. http://www.stats.govt.nz/Census/2013-census/profile-and-summary-reports/quickstats- about-national-highlights/cultural-diversity.aspx

van Dijk, J, Manchin, R, van Kesteren, J, Nevala, S & Hideg, G. 2007. *Methodology of the European Crime and Safety Survey 2005*. European Crime and Safety Survey Consortium. http://wp.unil.ch/icvs/files/2012/11/WP_methodology.pdf

Appendix A: Sampling

Interview cluster sizes

The distribution of interview cluster sizes for the overall sample is shown in Figure A1. The interview cluster sizes by the number of meshblocks for the main and Māori booster samples are provided in Table A1.

Figure A1: Distribution of interview cluster sizes, for the overall sample

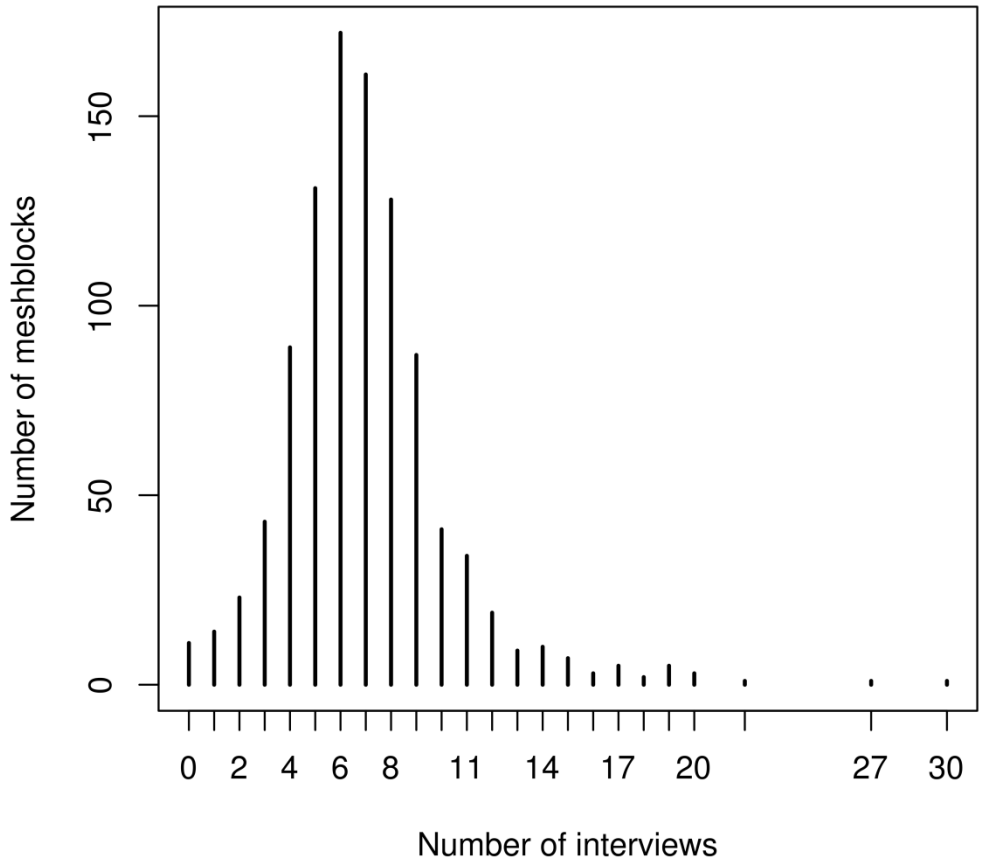


Table A1: Interview cluster sizes, by number of meshblocks for the main and Māori booster samples

Number of interviews (cluster size)	Main sample Number of meshblocks	Māori booster sample Number of meshblocks	Full sample Number of meshblocks
0	14	346	11
1	18	267	14
2	37	187	23
3	81	68	43
4	158	43	89
5	224	25	131
6	242	19	172
7	158	10	161
8	45	10	128
9	11	6	87
10	5	5	41
11	3	1	34
12	4	6	19
13	0	2	9
14	0	1	10
15	0	1	7
16	0	1	3
17	0	0	5
18	0	0	2
19	0	0	5
20 or more	0	2	6
Total	**1,000**	**1,000**	**1,000**
Average	5.2	1.7	6.9

The main sample was designed so that an average of 6.9 occupied private dwellings would be approached per meshblock, which under the targeted response rate of 70% would have resulted in an average of 4.8 main sample interviews per meshblock. The achieved response rate of 80% was higher than this target, resulting in an average of 5.2 main sample interviews per meshblock.

The Māori booster sample was not designed to achieve an average interview cluster size per meshblock; rather, it was designed to conduct a certain number of interviews (1660) given certain assumptions. This was higher than the 1,297 booster interviews conducted in NZCASS 2009, because the new booster sample design produces a higher design effect and therefore requires more booster sample interviews to achieve the same effective sample size. As discussed earlier, the achieved yield was lower than had been assumed, and more booster sample dwellings had to be selected in the last two fieldwork months to compensate. The final Māori booster sample contained 1708 interviews at an average of 1.7 Māori booster interviews per meshblock.

The average number of interviews per meshblock in the full sample is thus 6.9, compared to 6.1 in NZCASS 2009. Half of this increase was from the booster sample, as planned to maintain the same effective sample size. The other half was due to the higher than expected response rate in the main sample, and will have increased the effective sample size and thus improved the reliability of NZCASS 2014 estimates.

Total numbers of residential addresses compared with census counts

Residential addresses from the PAF were combined with addresses from the electoral roll where an elector of Māori descent resided to produce an initial sampling frame for the selection of dwellings in the 1000 meshblocks selected for the 2014 NZCASS. The total number of residential addresses in this combined database in the 1000 meshblocks was 64,379. In contrast, the total census count of occupied private dwellings in these meshblocks was 12.4% lower at 56,382.

The sampling database included many addresses that did not correspond to occupied private dwellings. Of the 7975 addresses selected as part of the main sample, 1165 (14.6%) were found to be 'Vacant/Not occupied', 'Not a dwelling', or an 'Empty section'. Taking this into account, the number of residential addresses in the sampling database seems slightly lower than desired.

Population growth since the 2013 Census also needs to be allowed for. The number of occupied private dwellings increased 6.1% between the 2006 Census and the 2013 Census, from 1,471,746 to 1,561,959 respectively (ie the average inter-censal growth rate was 0.9% per annum).[97] Fieldwork for the 2014 NZCASS took place from February to June 2014, approximately a year after the 2013 Census. Allowing for growth at the rate seen between the last two censuses, it is estimated that there would have been approximately 56,876 occupied private dwellings in the selected meshblocks during the survey period. Combining these adjustments to the totals indicates that the number of addresses in the sampling database is around 2227 (3.4%) lower than needed for full coverage.

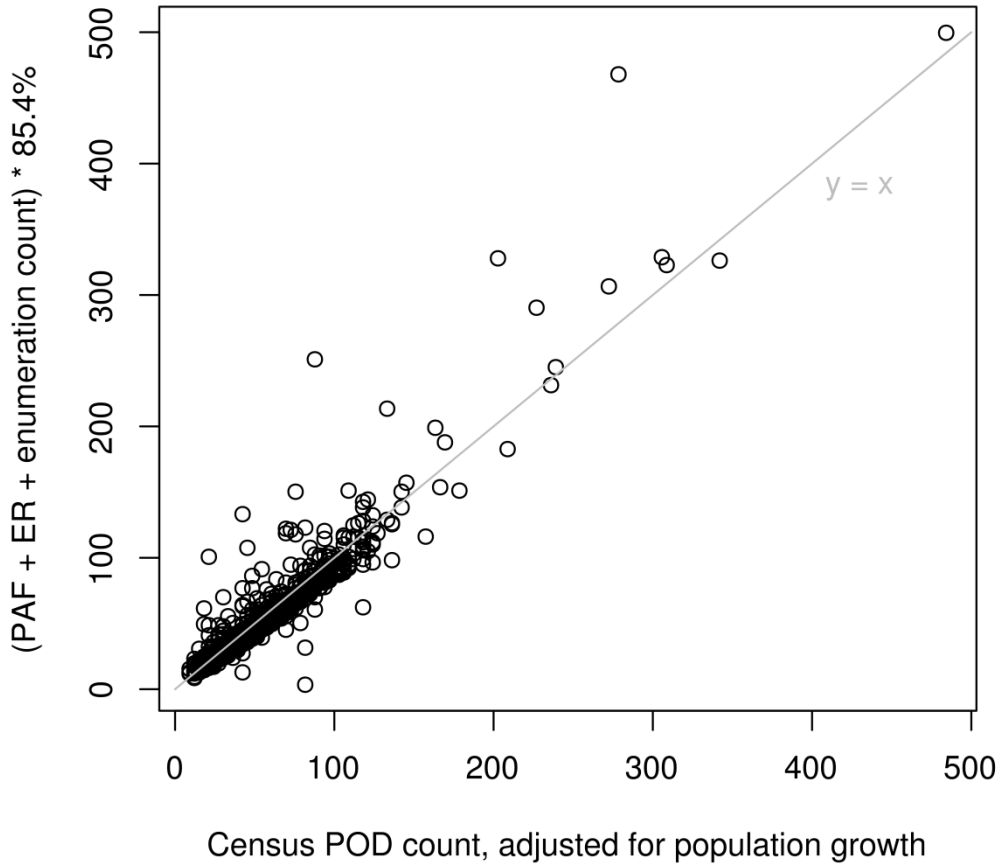
Interviewers were asked to list all dwellings in their assigned meshblock at addresses not in the sampling database (known as enumeration). A proportion of enumerated dwellings were selected to form part of the NZCASS sample. This enumeration process was intended to make up for any undercoverage in the sampling database. A total of 188 addresses were added to the database by enumeration (of which 19 were selected for interview), amounting to 8.4% of the estimated shortfall. This reduced the undercoverage to 3.1%, but did not eliminate it. This compares to an estimated undercoverage after enumeration of 1.6% for the 2009 NZCASS, as documented in Appendix A3 of the 2009 NZCASS Technical Report.

When a revised, larger booster sample was drawn for meshblocks scheduled for fieldwork in May and June, 367 addresses that were previously selected (and were not reselected) were unintentionally omitted from the sampling database provided to interviewers. Of these, 23 (6.3%) were enumerated, a rate that is consistent (allowing for random sampling variation) with the 8.4% estimate above for the whole sample.

[97] Restricting these totals to meshblocks containing 9 or more dwellings would give a closer approximation to the NZCASS sampling frame. However, this made very little difference to the inter-censal growth rates observed before the 2009 NZCASS (8.1% growth between 2001 and 2006, instead of 8.2%), so this has not been done here.

Undercoverage appeared to be fairly well spread, and not localised to a small proportion of meshblocks, although there was naturally some variation. Figure A2 plots the combined number of addresses from the PAF, electoral roll and enumeration (reduced by 14.6%) against the census counts adjusted for population growth.

Figure A2: Comparison of meshblock dwelling counts



ER = electoral roll
POD = private occupied dwellings

The median percentage shortfall was 7.9%, and the upper and lower quartiles were 0.6% and 12.5% respectively.

Number of adult occupants of each dwelling

One person aged 15 or over is interviewed for the NZCASS in each selected dwelling. This person is selected at random from a list (or roster) of all adult occupants gathered from the person first contacted at the dwelling. This gives people from large households a smaller probability of selection than people from small households, skewing the sample, so the data is weighted to compensate for this using the number of people aged 15 or over that usually live in that dwelling (or the number of Māori aged 15 or over, for booster sample selections). In booster sample dwellings, the person at the door was asked whether any of the adults listed in the roster might consider themselves Māori. The answer for each adult (Māori or Other) was recorded in the roster.

The number of adult occupants used for weighting is gathered later during the interview, however, and can differ from the number on the roster produced at the initial contact. In previous NZCASS iterations, occupant rostering and respondent selection was conducted on forms separate from the questionnaire, and was not available for weighting purposes. In the 2014 NZCASS, the household roster data is entered directly and can be readily linked to data gathered in the interview with the ultimate respondent. It was decided to still gather the number of eligible household residents from the respondent during the questionnaire and use this for weighting purposes in the 2014 NZCASS, for consistency with previous iterations. However, it is now possible to investigate the differences between the numbers of adult occupants gathered via these two different methods. The overall percentage of dwellings where these counts agree was mentioned earlier in the 'occupancy match rate' in Table 5.13: Table A2 presents more detail.

Table A2: Comparison of number of adult occupants from two different sources

Number of usual residents in dwelling aged 15 or over										
	From responses to Q146A and Q147									
Initial roster	1	2	3	4	5	6	7	8	10	Total
1	1,815	73	17	8	2					1,915
2	37	3,306	95	25	8					3,471
3	14	44	795	49	11	3				916
4	3	17	22	378	26	4	2	1		453
5	1	4	5	8	103	6	2		1	130
6		3	2	1	4	19	2			31
7			1			5	11	1		18
8			2					3		5
10					2				1	3
12						1				1
Total	**1,870**	**3,447**	**939**	**469**	**156**	**38**	**17**	**5**	**2**	**6,943**

The two counts agree for 92.6% of respondents. Agreement is higher in smaller households (97% among respondents reporting 1 adult occupant and 95% among respondents reporting 2 adult occupants), and lower in larger households (85% among respondents reporting 3 adult occupants, 81% among respondents reporting 4 adult occupants, and 63% among respondents reporting 5 or more adult occupants). While discrepancies go both ways, it is more common for the respondent's count to be higher (336 cases) than vice versa (176 cases).

There could be several reasons for these differences, such as confusion about whether children should be included, or different views on who usually lives there. The main implication for the NZCASS is that weighting using respondent-reported occupant counts will not have adjusted perfectly for the respondent selection process, making the survey results less precise. In future iterations, weighting using counts from the initial household occupant roster would be worth considering, although this may have implications for comparability with previous surveys.

Household and respondent declines by demographics

The demographics of the households and respondents who declined to participate in the NZCASS are included in tables A3 to A6. The number of declines is presented for combinations of level of area deprivation, 2013 Police crime rate area and gender.

Table A3: Household and respondent declines by deprivation

	Household decline			Respondent decline			
Level of area deprivation (NZDep2013)	Main	Booster	Total	Main	Booster	Total	Grand total
1 (lowest)	89	16	105	32	5	37	**142**
2	93	21	114	20	2	22	**136**
3	82	32	114	25	7	32	**146**
4	96	30	126	22	13	35	**161**
5 (highest)	83	44	127	18	18	36	**163**
NA	2		2				**2**
Grand total	**445**	**143**	**588**	**117**	**45**	**162**	**750**

Table A4: Household and respondent declines by Police recorded crime groups

	Household decline			Respondent decline			
Crime rate (2013)[98]	Main	Booster	Total	Main	Booster	Total	Grand total
Low	165	24	189	46	7	53	**242**
Medium	126	35	161	34	10	44	**205**
High	154	84	238	37	28	65	**303**
Grand total	**445**	**143**	**588**	**117**	**45**	**162**	**750**

[98] Crime rate groups have been derived for each meshblock from Police recorded crime data. The Police data relates to the number of incidents recorded as crimes (in -scope in NZCASS) that occurred in 2013. This produces frequency counts per meshblock, which are grouped into three evenly sized groups weighted by the 2013 estimated resident population 2013.

Table A5: Respondent declines by gender and deprivation

	Main sample			Māori booster sample			
Level of area deprivation (NZDep2013)	Female	Male	Total	Female	Male	Total	Grand total
1 (lowest)	15	17	32	3	2	5	**37**
2	11	9	20	2	0	2	**22**
3	14	11	25	5	2	7	**32**
4	12	10	22	5	8	13	**35**
5 (highest)	11	7	18	8	10	18	**36**
Grand total	**63**	**54**	**117**	**23**	**22**	**45**	**162**

Table A6: Respondent declines by gender and Police recorded crime groups

	Main sample			Māori booster sample			
Crime Rate (2013)	Female	Male	Total	Female	Male	Total	Grand Total
Low	27	19	46	3	4	7	**53**
Medium	13	21	34	3	7	10	**44**
High	23	14	37	17	11	28	**65**
Grand Total	**63**	**54**	**117**	**23**	**22**	**45**	**162**

Appendix B: Fieldwork products

Letter to household

<Date>

<Address 1>
<Address 2>
<Address 3>

Dear Householder

The Ministry of Justice is carrying out a survey about crime and safety in New Zealand and we invite you to take part.

We do this to gather important information to help a range of government and non-government organisations to develop better services for people affected by crime and create safer communities.

By talking to New Zealanders we are able to better understand how much crime occurs, what is reported, who experiences crime and how people are affected by it.

The survey is being undertaken for the Ministry by CBG Public Sector Surveying. Your household has been randomly selected to take part and in the next few weeks an interviewer [**name in here**] wearing an identification badge will call at your address.

The interviewer may invite someone from your household who is 15 years or older to do the survey. The survey is completely confidential. Your participation in the survey is completely voluntary.

The interviewer will explain more about the survey and answer any questions.

If you would like more information then please call CBG Public Sector Surveying on **0800 478 783** or the Ministry of Justice on **0800 464 656.**

We rely on the goodwill of those invited to take part to make the survey a success and hope you will take part.

Thank you in advance.

Andrew Bridgman
Chief Executive and Secretary for Justice

Information pamphlet

We'd like to find out more about New Zealanders' feelings of safety and their experience of crime.

This will inform the work being done by a range of agencies to create safer neighbourhoods and communities.

This survey is carried out nationwide every few years.

Please help us by sharing your views and experiences.

Thank you very much for helping us with this survey.

More information

Where can I find out more information about this survey?

CBG Public Sector Surveying (CBG)
0800 478 783 tollfree

Ministry of Justice
0800 464 656 tollfree

Who can I call for support if I have been a victim of crime?

You can call the Victims of Crime Information Line tollfree on 0800 650 654 or visit www.victimsinfo.govt.nz

If I have been a victim of crime and I want to report it, who should I contact?

You can call or visit your local police station or call the anonymous Crimestoppers number 0800 555 111.

In the case of an emergency, call 111.

MINISTRY OF JUSTICE
Tāhū o te Ture

CBG PUBLIC SECTOR SURVEYING

New Zealand Government

MOJ0116J-JAN13

MINISTRY OF JUSTICE
Tāhū o te Ture

New Zealand **Crime & Safety Survey** 2014

What is the New Zealand Crime and Safety Survey?

The New Zealand Crime and Safety Survey collects information about New Zealanders' feelings of safety and their experience of crime. We carried out similar surveys in 1996, 2001, 2006 and 2009.

About 6000 people in households throughout the country will be taking part in the survey.

What questions will be asked?

The questions are about:

- how safe you feel in your community and how this might affect your everyday life
- if you have experienced any crimes and, if so, what the effects were
- whether or not you have told anyone about them
- how helpful any agencies were (if they were contacted).

Who is carrying out the survey?

The survey is being carried out by CBG Public Sector Surveying (CBG) on behalf of the Ministry of Justice.

CBG is an independent, New Zealand-based research company.

Who will be asked to take part?

One person from your household aged 15 or over may be randomly selected to take part in the survey. You don't need to have been a victim of crime to answer the questions. Your participation in the survey is voluntary.

How will the questions be asked?

The interviewers will use laptops rather than paper questionnaires. If there is anything you don't want to talk about, you can type it into the laptop yourself.

How long will the interview take?

The interview length can vary depending on how often you have been a victim of crime. It is likely to be around 40 minutes if you have not experienced any crime. It will take longer if you have been the victim of crime recently. The interview can be done at a date and time that suits you.

What will the information be used for?

The results from the survey will inform work being done by a range of justice sector agencies to create safer neighbourhoods and communities. Without the survey, there would be a big gap in the knowledge base on New Zealanders' experiences.

How is my privacy protected?

The information you provide to the interviewer is confidential and protected by the Privacy Act 1993.

This means the interviewer will not discuss your information with anyone else. The answers you give are added to other people's answers to create group statistics. The statistics, which are coded for computer analysis, will be kept and may be used by approved researchers.

Your name and identifying details will not be included in the published material.

Where can I find the survey results?

The results of the main survey will be published on the Ministry of Justice website justice.govt.nz

It is expected the key findings will be released to the public in 2015.

Thank-you card

MINISTRY OF JUSTICE
Tāhū o te Ture

CBG
PUBLIC SECTOR SURVEYING

Thank you for taking part in the New Zealand Crime and Safety Survey 2014

Dear survey participant

Thank you very much for participating in the New Zealand Crime and Safety Survey 2014. Information from this and previous surveys is used to support work by justice sector agencies to create safer neighbourhoods and communities. We appreciate your time and effort.

If you are affected by crime in the future call the police. Information and support is available. Call the Victims of Crime Information Line 0800 650 654, or one of the agencies listed on the right.

To report a crime, you can call or visit your local police station or call the anonymous Crimestoppers number 0800 555 111.

In the case of an emergency, **call 111**.

For more information about the survey, please call CBG Public Sector Surveying 0800 478 783 or the Ministry of Justice 0800 464 656.

Listed here are some of the agencies that provide support for victims of crime:

Victim Support
0800 VICTIM
0800 842 846

Family Violence
Information Line
0800 456 450

Lifeline
0800 54 33 54

Youthline
0800 37 66 33

Women's Refuge
0800 REFUGE
0800 733 843

Victims Information
victimsinfo.govt.nz
0800 650 654

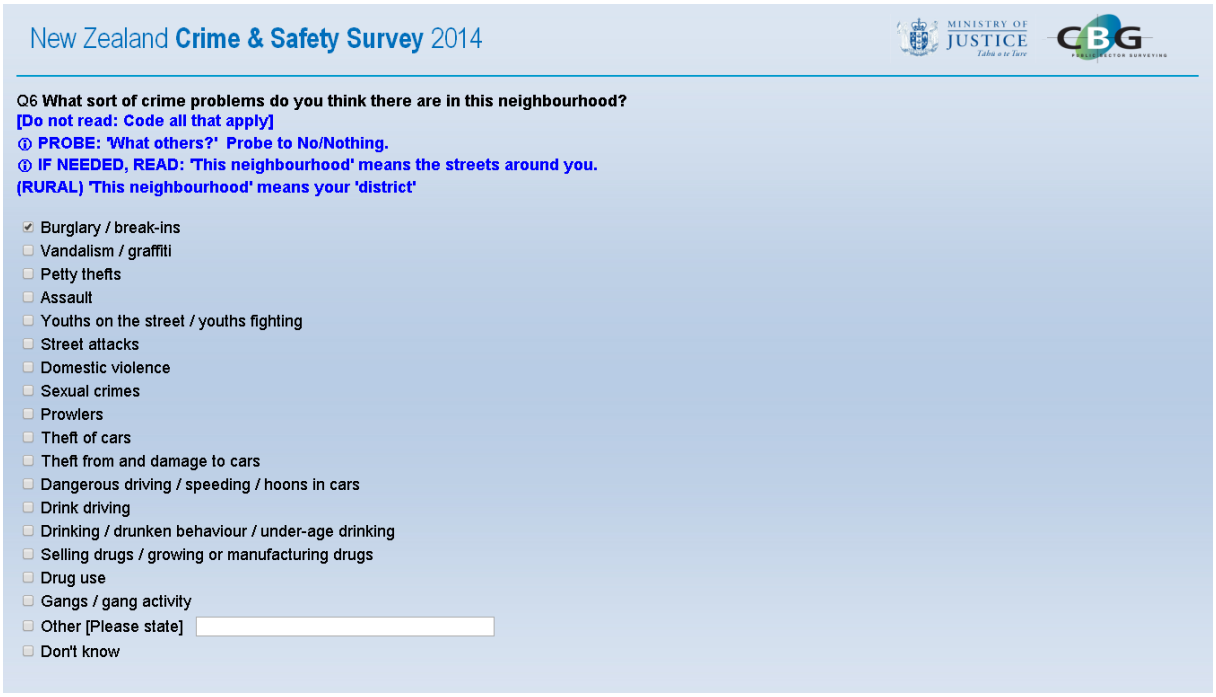
For the following services, please call the Victims of Crime Information Line on 0800 650 654 for the local number in your area:

Age Concern
Rape Crisis

MOJ0116.2-DEC13

Appendix C: Questionnaire screenshots

The following screenshots aim to demonstrate the look and feel of the questionnaire in 2014. These screenshots also provide a visual reference during questionnaire programming stages of future surveys.



New Zealand **Crime & Safety Survey** 2014

MINISTRY OF JUSTICE Tāhū o te Ture

Using the categories on Showcard B, can you tell me how much of a problem you think the following things are in your neighbourhood?
ⓘ PROBE, IF NEEDED, READ: 'This neighbourhood' means the streets around you.
(RURAL) 'This neighbourhood' means your 'district'

[Showcard B, page 2]
Q8.5 **Drunks, glue sniffers or people high on drugs on the streets.**

- 1. A very big problem
- 2. A fairly big problem
- 3. Not a very big problem
- 4. Not a problem at all
- 98. Don't know

⇦ Back | Continue ⇨ | Reset these answers

New Zealand **Crime & Safety Survey** 2014

MINISTRY OF JUSTICE Tāhū o te Ture

VICTIMISATION SCREENER QUESTIONS

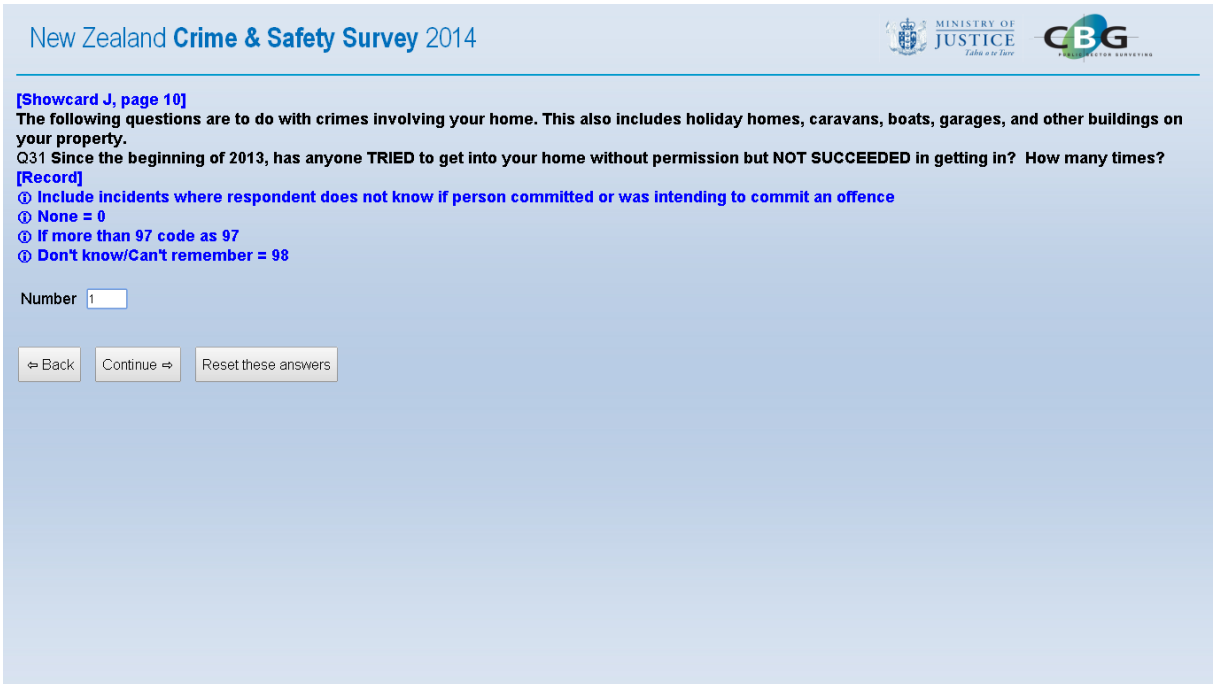
ⓘ Use Showcard J, page 10. Leave showcard in view for screeners i.e. till Q44

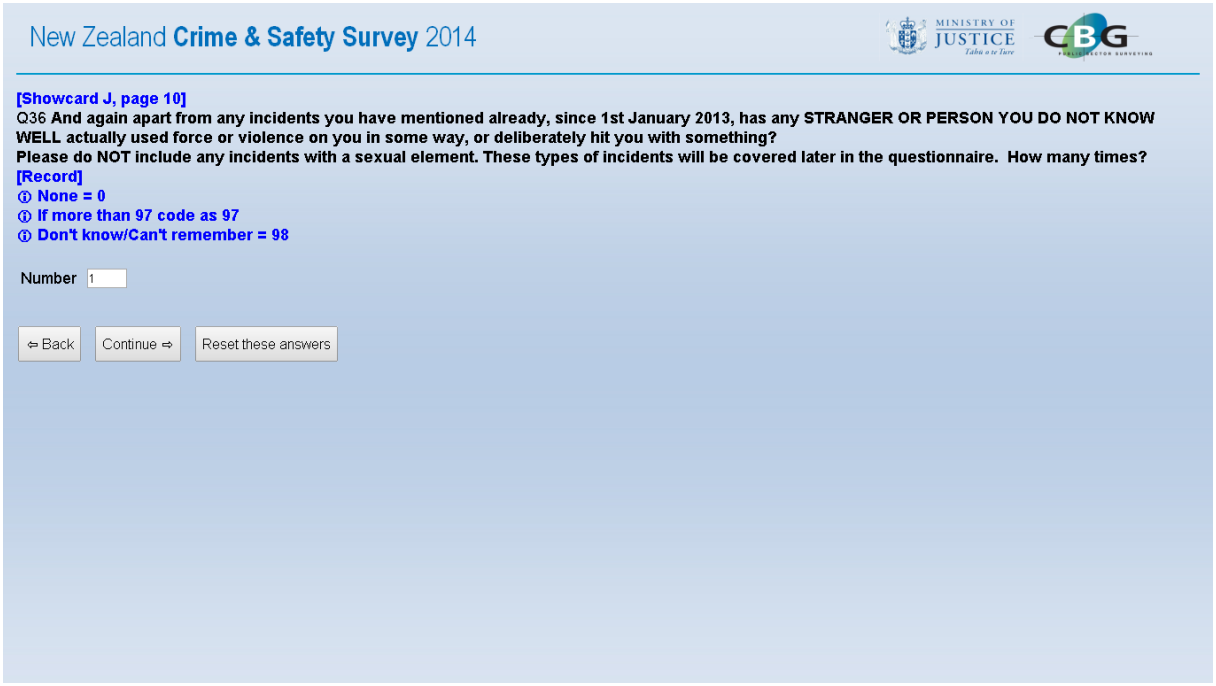
I'd now like to ask you about some things that might have happened to you or your household since 1st January 2013, that is, since the beginning of last year, in which you may have been the victim of a crime or offence. This doesn't mean that crimes that may have happened before this time are unimportant, but we want to get the most recent information on people's experience of crime.
Showcard J lists the sorts of incidents we are interested in. These...
- Must have happened since 1st JANUARY 2013
- Must have HAPPENED TO YOU PERSONALLY, or to YOUR HOUSEHOLD PROPERTY
- Must have happened in NEW ZEALAND
- Can be both SERIOUS things and SMALL things too.
It is often difficult to remember exactly when things happen, so take the time you need.

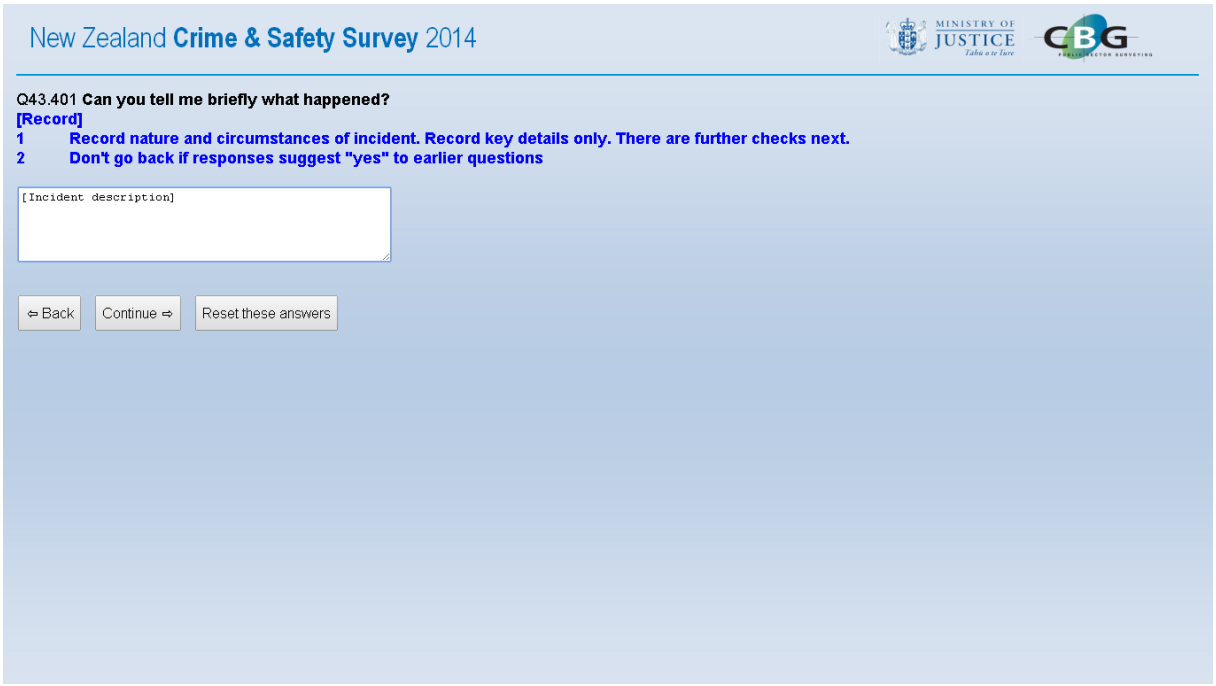
[Showcard J, page 10]
Q27 **First, has anyone NOW in this household owned or had the regular use of a car, motorcycle, van or truck at any time since 1st January 2013?**
ⓘ Code yes if anyone has owned or had regular use of a car, motorcycle, van or truck since the 1st January 2013, but no longer does

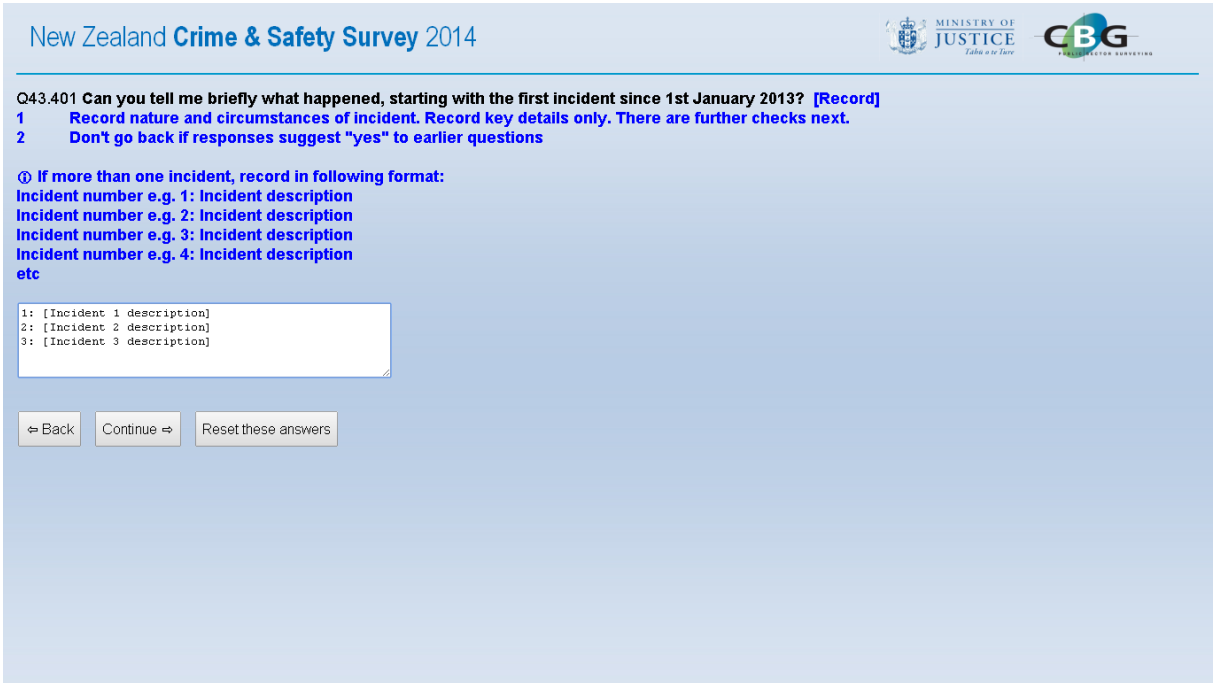
- Yes
- No

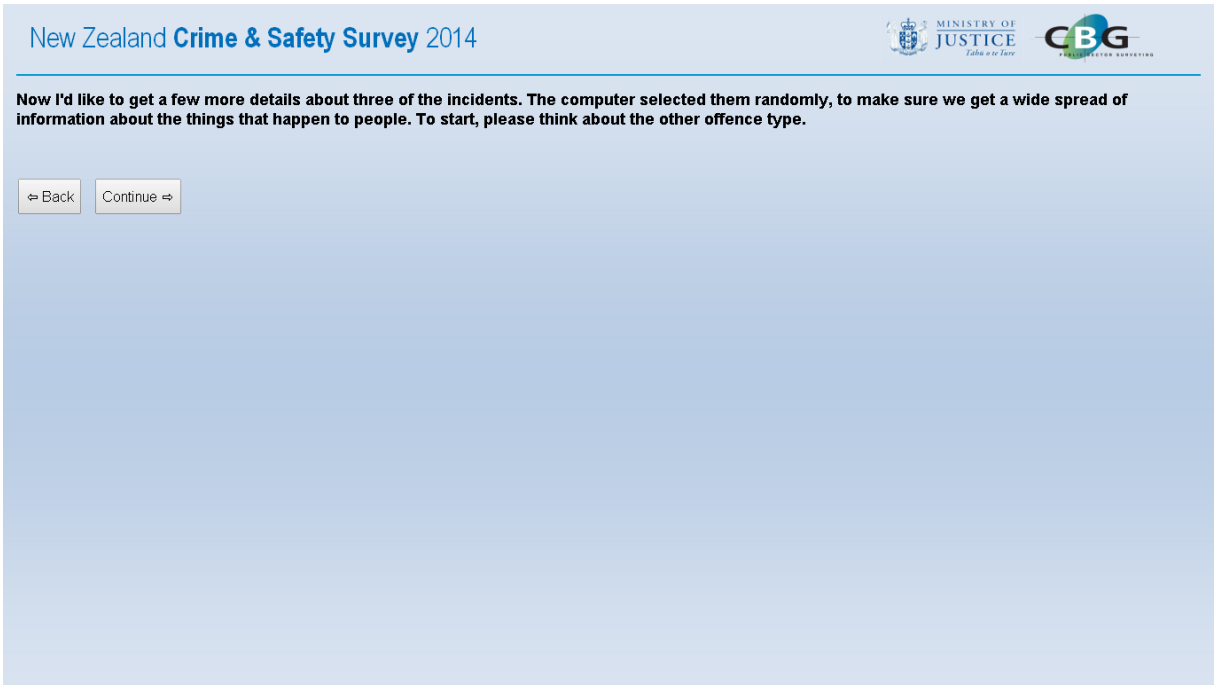
⇦ Back | Continue ⇨ | Reset these answers

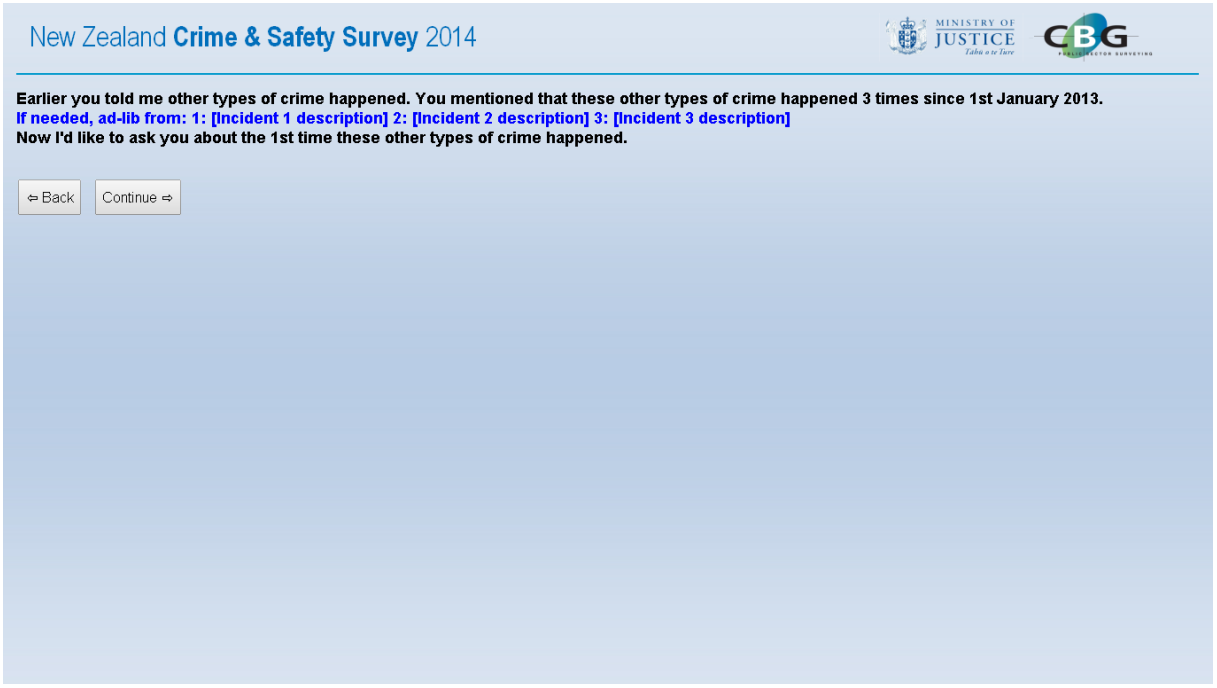

New Zealand Crime & Safety Survey 2014
MINISTRY OF JUSTICE
[Showcard J, page 10]
The following questions are to do with crimes involving your home. This also includes holiday homes, caravans, boats, garages, and other buildings on your property.
Q31 Since the beginning of 2013, has anyone TRIED to get into your home without permission but NOT SUCCEEDED in getting in? How many times?
[Record]
Include incidents where respondent does not know if person committed or was intending to commit an offence
None = 0
If more than 97 code as 97
Don't know/Can't remember = 98
Number 1
⇦ Back
Continue ⇨
Reset these answers

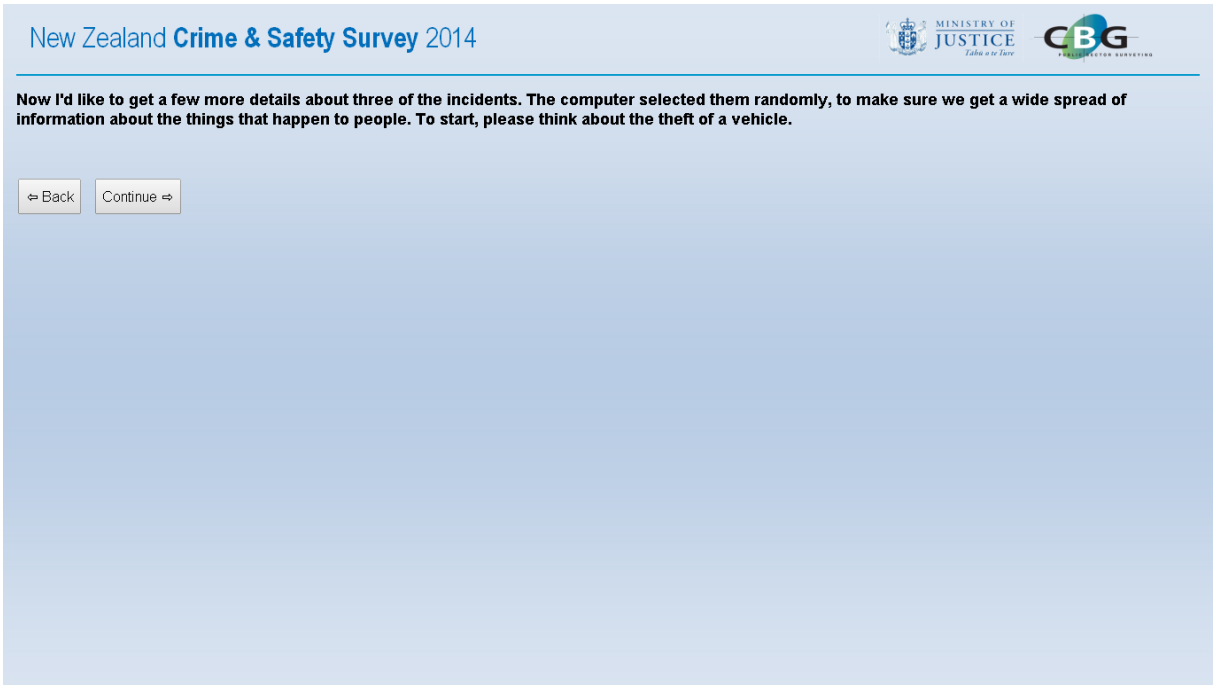

New Zealand Crime & Safety Survey 2014
MINISTRY OF JUSTICE
[Showcard J, page 10]
Q36 And again apart from any incidents you have mentioned already, since 1st January 2013, has any STRANGER OR PERSON YOU DO NOT KNOW WELL actually used force or violence on you in some way, or deliberately hit you with something?
Please do NOT include any incidents with a sexual element. These types of incidents will be covered later in the questionnaire. How many times?
[Record]
None = 0
If more than 97 code as 97
Don't know/Can't remember = 98
Number 1
⇦ Back
Continue ⇨
Reset these answers


New Zealand Crime & Safety Survey 2014
MINISTRY OF JUSTICE
Q43.401 Can you tell me briefly what happened?
[Record]
1 Record nature and circumstances of incident. Record key details only. There are further checks next.
2 Don't go back if responses suggest "yes" to earlier questions
[Incident description]
⇦ Back
Continue ⇨
Reset these answers

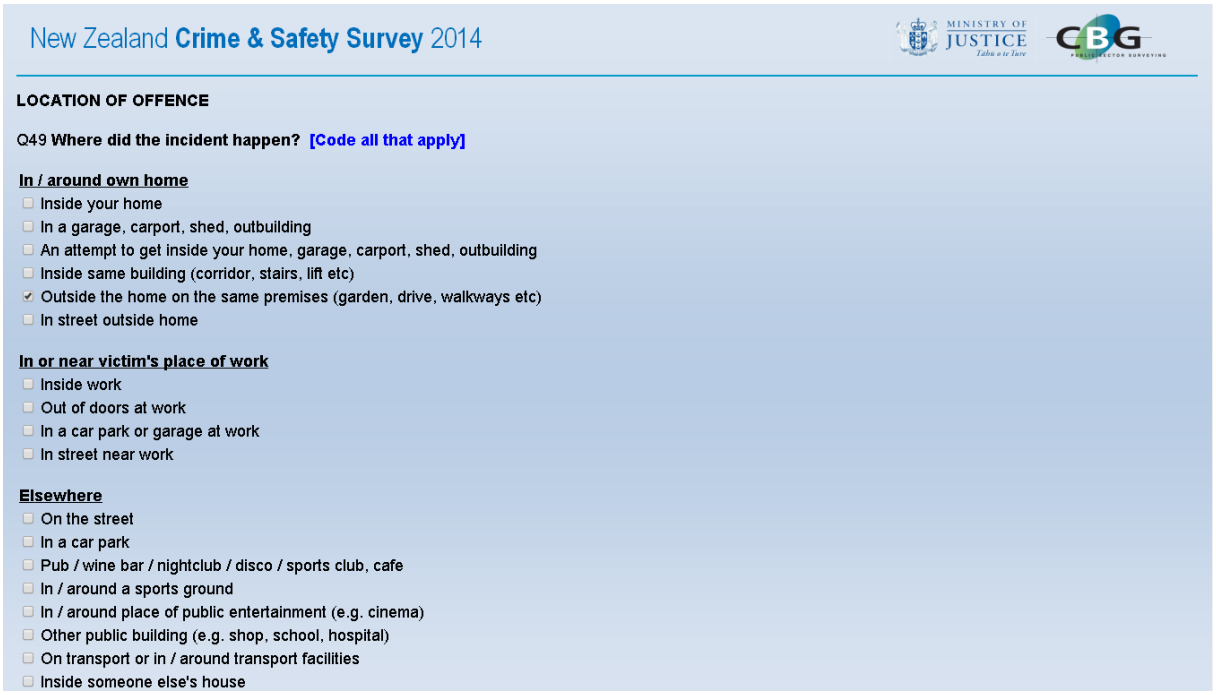

New Zealand Crime & Safety Survey 2014
MINISTRY OF JUSTICE
Q43.401 Can you tell me briefly what happened, starting with the first incident since 1st January 2013? [Record]
1 Record nature and circumstances of incident. Record key details only. There are further checks next.
2 Don't go back if responses suggest "yes" to earlier questions
ⓘ If more than one incident, record in following format:
Incident number e.g. 1: Incident description
Incident number e.g. 2: Incident description
Incident number e.g. 3: Incident description
Incident number e.g. 4: Incident description
etc
1: [Incident 1 description]
2: [Incident 2 description]
3: [Incident 3 description]
⇦ Back
Continue ⇨
Reset these answers


New Zealand Crime & Safety Survey 2014
MINISTRY OF JUSTICE
Now I'd like to get a few more details about three of the incidents. The computer selected them randomly, to make sure we get a wide spread of information about the things that happen to people. To start, please think about the other offence type.
⇦ Back
Continue ⇨


New Zealand Crime & Safety Survey 2014
MINISTRY OF JUSTICE
Earlier you told me other types of crime happened. You mentioned that these other types of crime happened 3 times since 1st January 2013.
If needed, ad-lib from: 1: [Incident 1 description] 2: [Incident 2 description] 3: [Incident 3 description]
Now I'd like to ask you about the 1st time these other types of crime happened.
⇦ Back
Continue ⇨


New Zealand Crime & Safety Survey 2014
MINISTRY OF JUSTICE
Now I'd like to get a few more details about three of the incidents. The computer selected them randomly, to make sure we get a wide spread of information about the things that happen to people. To start, please think about the theft of a vehicle.
⇦ Back
Continue ⇨


New Zealand Crime & Safety Survey 2014
MINISTRY OF JUSTICE
DATE OF OFFENCE
Q44 Can I just confirm that this incident happened in New Zealand AND after 1st January 2013?
Yes
No
⇦ Back
Continue ⇨
Reset these answers



New Zealand **Crime & Safety Survey** 2014

MINISTRY OF JUSTICE

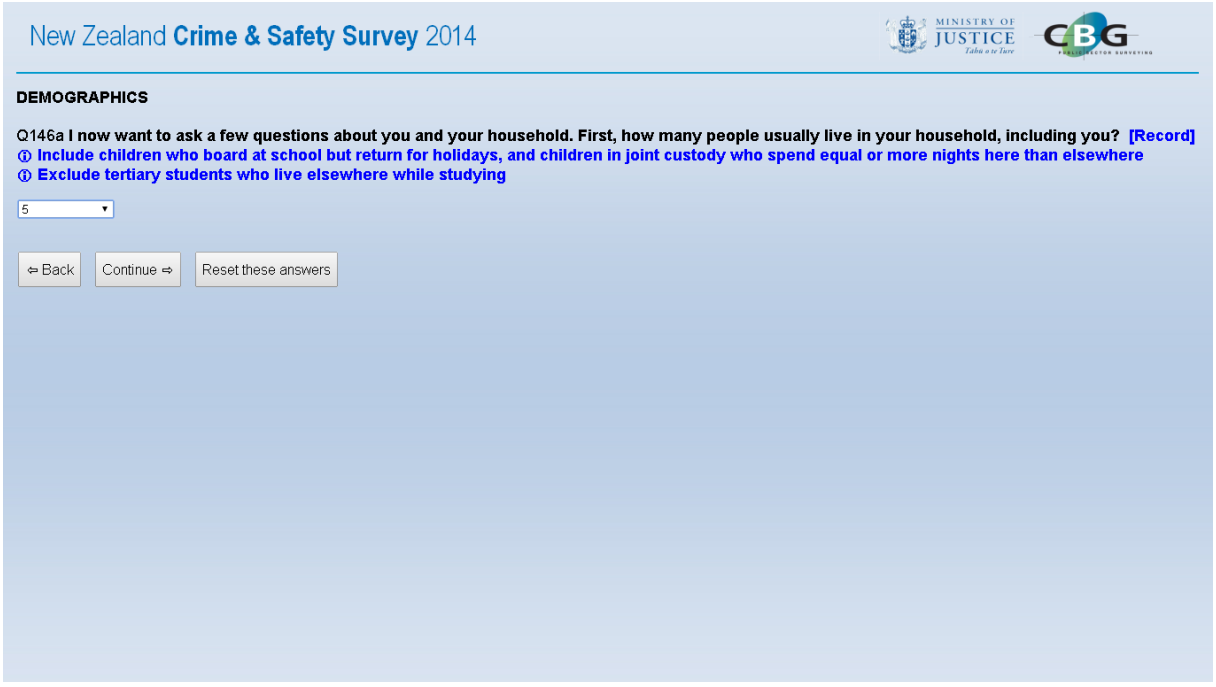
LOCATION OF OFFENCE

Q49 **Where did the incident happen?** **[Code all that apply]**

In / around own home
- ☐ Inside your home
- ☐ In a garage, carport, shed, outbuilding
- ☐ An attempt to get inside your home, garage, carport, shed, outbuilding
- ☐ Inside same building (corridor, stairs, lift etc)
- ☑ Outside the home on the same premises (garden, drive, walkways etc)
- ☐ In street outside home

In or near victim's place of work
- ☐ Inside work
- ☐ Out of doors at work
- ☐ In a car park or garage at work
- ☐ In street near work

Elsewhere
- ☐ On the street
- ☐ In a car park
- ☐ Pub / wine bar / nightclub / disco / sports club, cafe
- ☐ In / around a sports ground
- ☐ In / around place of public entertainment (e.g. cinema)
- ☐ Other public building (e.g. shop, school, hospital)
- ☐ On transport or in / around transport facilities
- ☐ Inside someone else's house



New Zealand **Crime & Safety Survey** 2014

MINISTRY OF JUSTICE

DEMOGRAPHICS

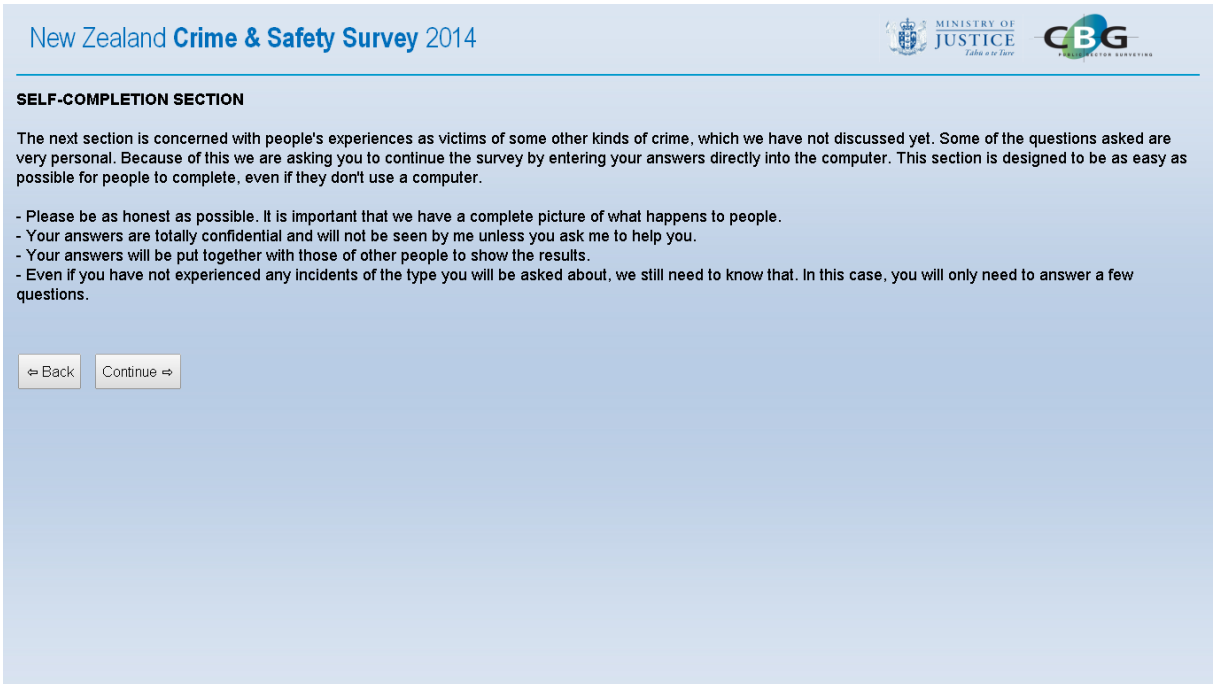
Q146a **I now want to ask a few questions about you and your household. First, how many people usually live in your household, including you?** **[Record]**
ⓘ **Include children who board at school but return for holidays, and children in joint custody who spend equal or more nights here than elsewhere**
ⓘ **Exclude tertiary students who live elsewhere while studying**

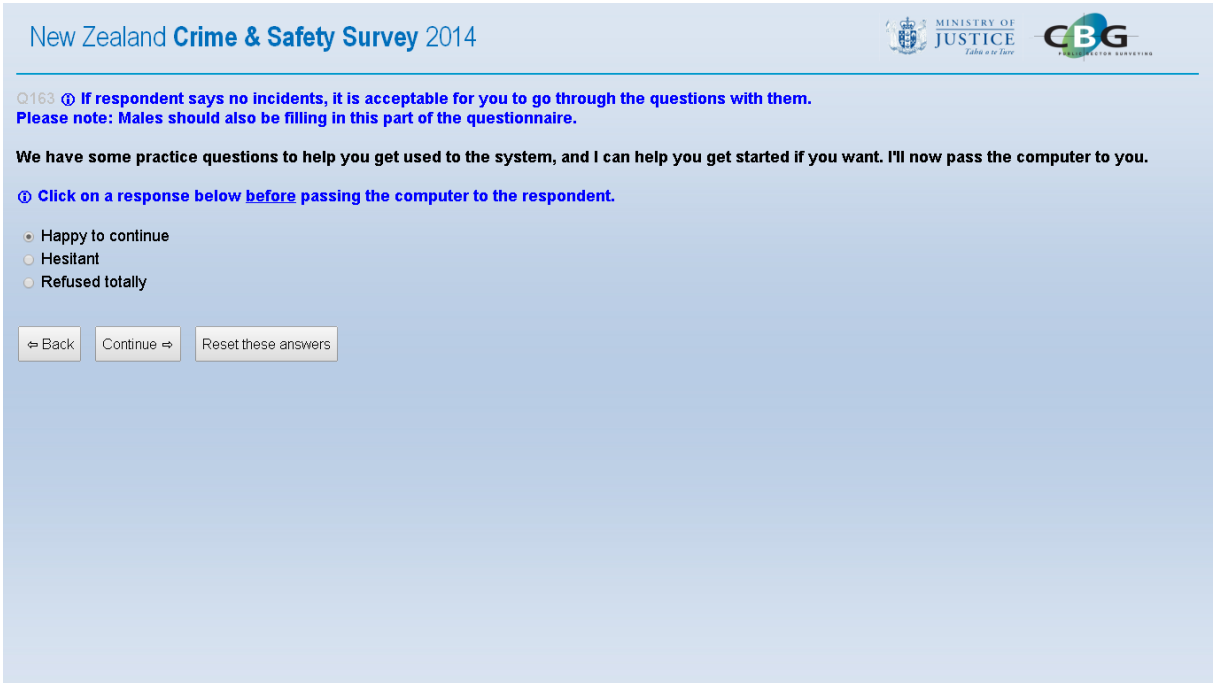
5

⇦ Back | Continue ⇨ | Reset these answers

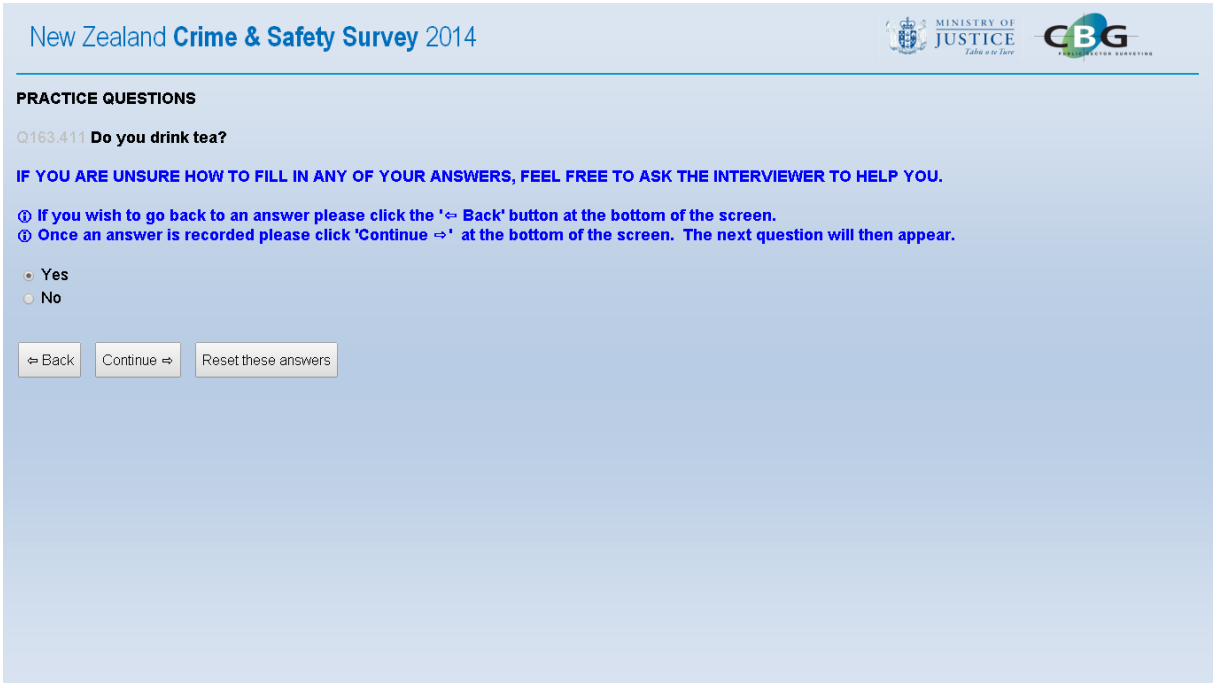

New Zealand Crime & Safety Survey 2014
MINISTRY OF JUSTICE
[Showcard Y, page 25]
Q151 Please use this card to tell me which ethnic group or groups you belong to. [Code all that apply]
1. New Zealand European
2. Maori
3. Samoan
4. Cook Island Maori
5. Tongan
6. Niuean
7. Chinese
8. Indian
96. Other (such as Dutch, Japanese, Tokelauan) [Please state]
99. Refused
⇦ Back
Continue ⇨
Reset these answers

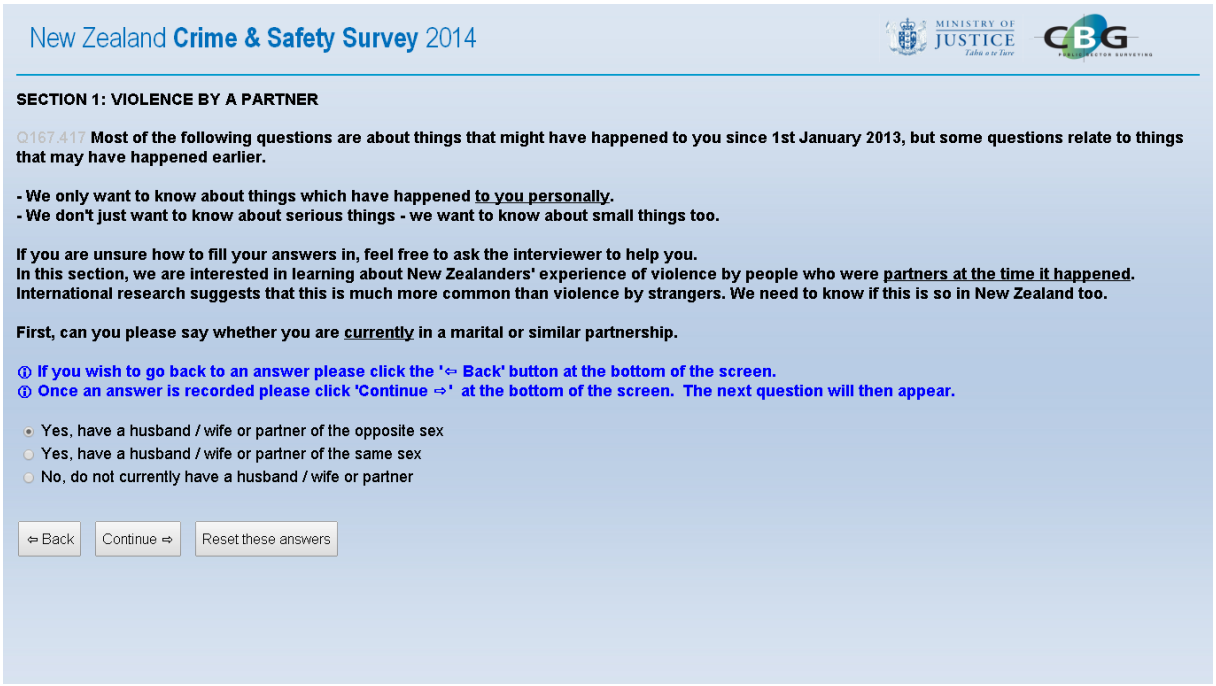

New Zealand Crime & Safety Survey 2014
MINISTRY OF JUSTICE
Q151_Other1
ai
aitutaki islander
Q151_Other2
Q151_Other3
Q151_Other4
Q151_Other5
Q151_Other6
⇦ Back
Continue ⇨
Reset these answers


New Zealand Crime & Safety Survey 2014
MINISTRY OF JUSTICE
SELF-COMPLETION SECTION
The next section is concerned with people's experiences as victims of some other kinds of crime, which we have not discussed yet. Some of the questions asked are very personal. Because of this we are asking you to continue the survey by entering your answers directly into the computer. This section is designed to be as easy as possible for people to complete, even if they don't use a computer.
- Please be as honest as possible. It is important that we have a complete picture of what happens to people.
- Your answers are totally confidential and will not be seen by me unless you ask me to help you.
- Your answers will be put together with those of other people to show the results.
- Even if you have not experienced any incidents of the type you will be asked about, we still need to know that. In this case, you will only need to answer a few questions.
⇦ Back
Continue ⇨


New Zealand Crime & Safety Survey 2014
MINISTRY OF JUSTICE
Q163 ⓘ If respondent says no incidents, it is acceptable for you to go through the questions with them.
Please note: Males should also be filling in this part of the questionnaire.
We have some practice questions to help you get used to the system, and I can help you get started if you want. I'll now pass the computer to you.
ⓘ Click on a response below before passing the computer to the respondent.
Happy to continue
Hesitant
Refused totally
⇦ Back
Continue ⇨
Reset these answers


New Zealand Crime & Safety Survey 2014
MINISTRY OF JUSTICE
CBG
PRACTICE QUESTIONS
Q163.101
Continue with practice questions
Continue without practice questions
⇦ Back
Continue ⇨
Reset these answers


New Zealand Crime & Safety Survey 2014
MINISTRY OF JUSTICE
CBG
PRACTICE QUESTIONS
Q163.411 Do you drink tea?
IF YOU ARE UNSURE HOW TO FILL IN ANY OF YOUR ANSWERS, FEEL FREE TO ASK THE INTERVIEWER TO HELP YOU.
If you wish to go back to an answer please click the '⇦ Back' button at the bottom of the screen.
Once an answer is recorded please click 'Continue ⇨' at the bottom of the screen. The next question will then appear.
Yes
No
⇦ Back
Continue ⇨
Reset these answers



New Zealand **Crime & Safety Survey** 2014

MINISTRY OF JUSTICE Tāhū o te Ture

CBG PUBLIC SECTOR SURVEYING

SECTION 1: VIOLENCE BY A PARTNER

Q167.417 **Most of the following questions are about things that might have happened to you since 1st January 2013, but some questions relate to things that may have happened earlier.**

- We only want to know about things which have happened to you personally.
- We don't just want to know about serious things - we want to know about small things too.

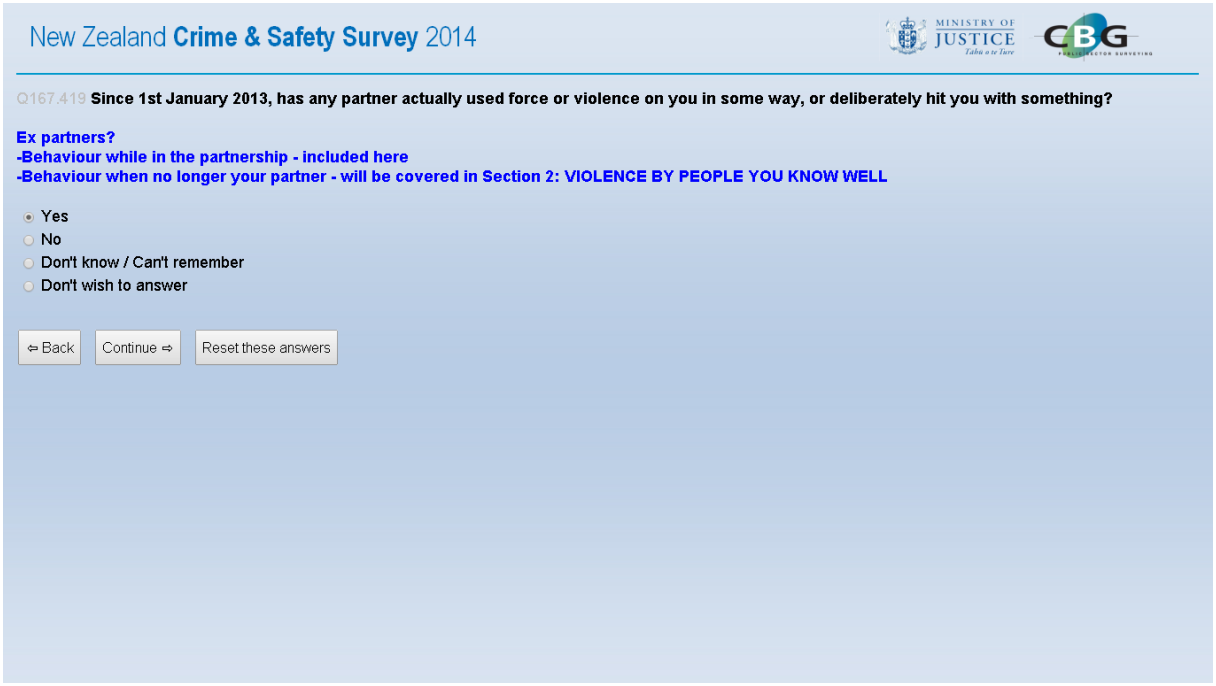
If you are unsure how to fill your answers in, feel free to ask the interviewer to help you.
In this section, we are interested in learning about New Zealanders' experience of violence by people who were partners at the time it happened. International research suggests that this is much more common than violence by strangers. We need to know if this is so in New Zealand too.

First, can you please say whether you are currently in a marital or similar partnership.

ⓘ If you wish to go back to an answer please click the '⇦ Back' button at the bottom of the screen.
ⓘ Once an answer is recorded please click 'Continue ⇨' at the bottom of the screen. The next question will then appear.

- (•) Yes, have a husband / wife or partner of the opposite sex
- ( ) Yes, have a husband / wife or partner of the same sex
- ( ) No, do not currently have a husband / wife or partner

⇦ Back | Continue ⇨ | Reset these answers



New Zealand **Crime & Safety Survey** 2014

MINISTRY OF JUSTICE Tāhū o te Ture

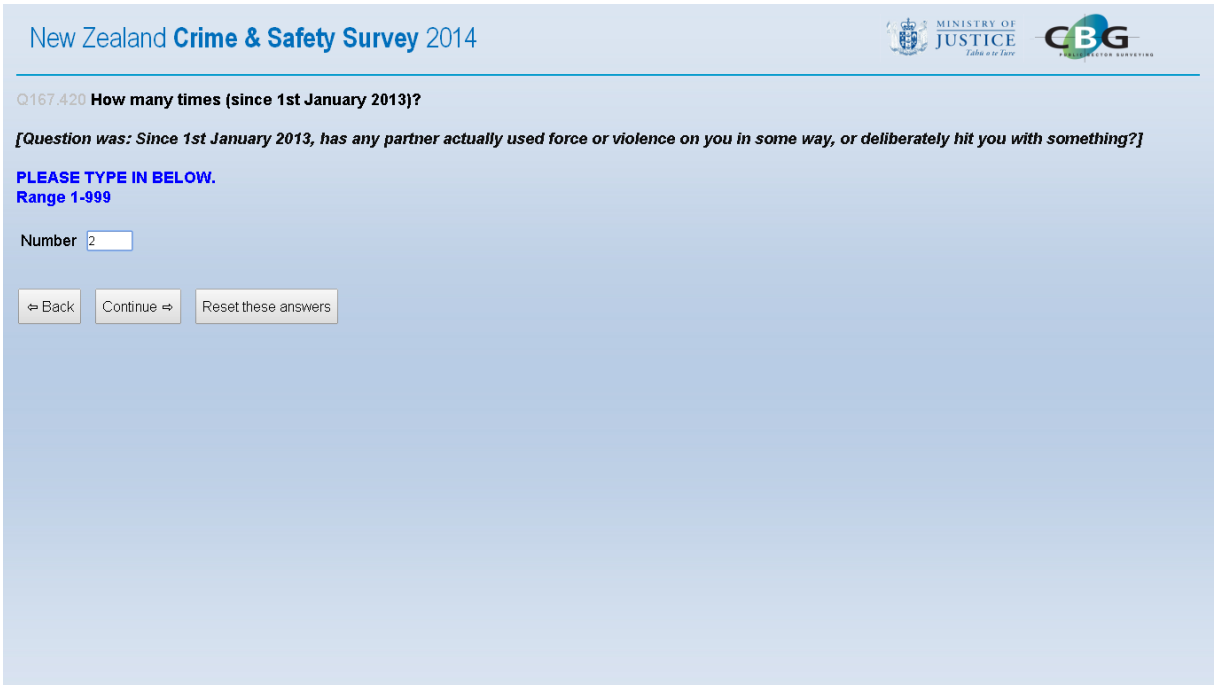
CBG PUBLIC SECTOR SURVEYING

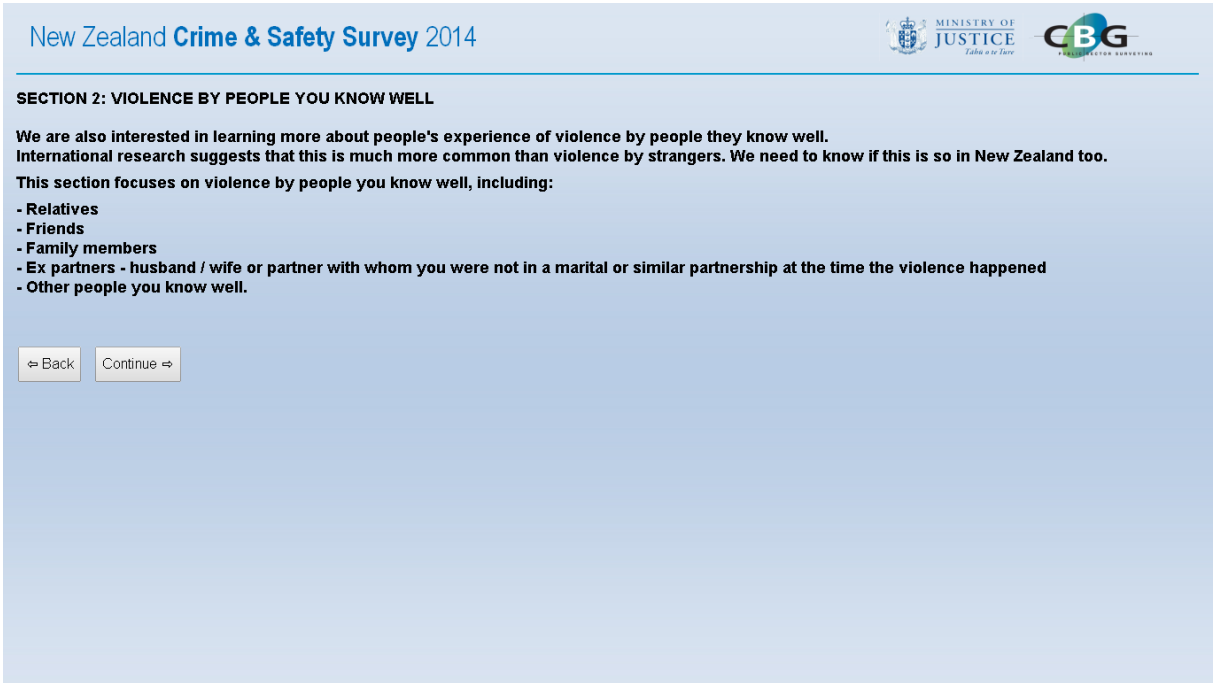
Q167.419 **Since 1st January 2013, has any partner actually used force or violence on you in some way, or deliberately hit you with something?**

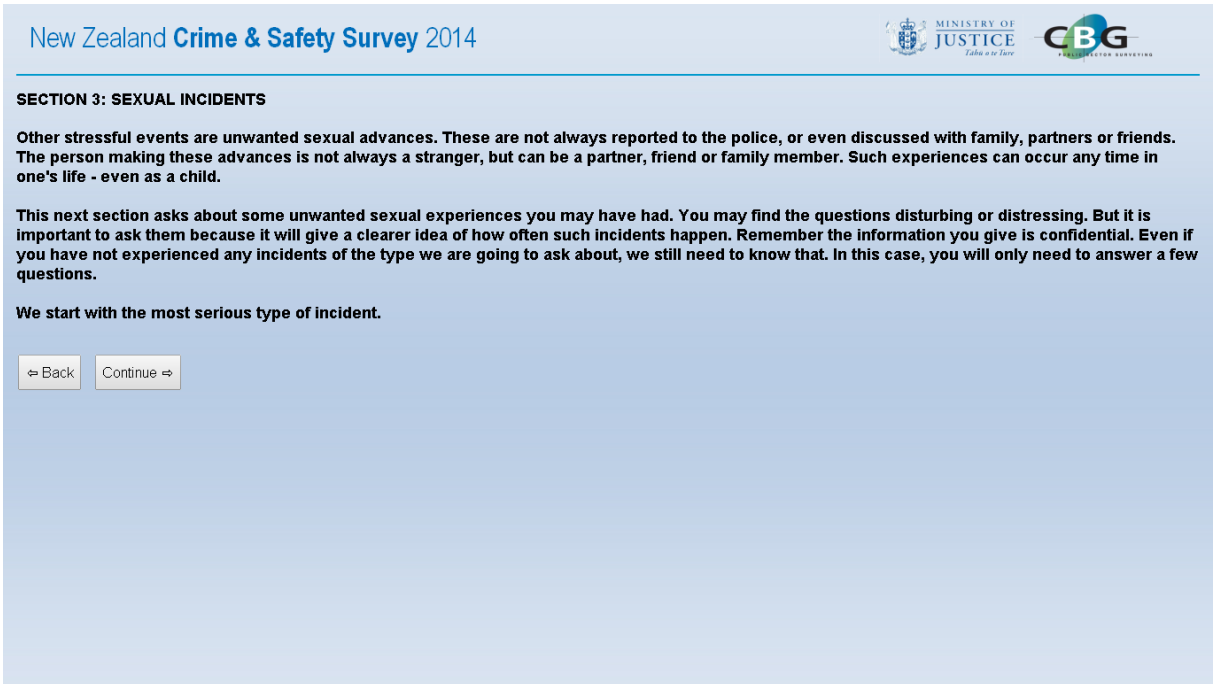
Ex partners?
-Behaviour while in the partnership - included here
-Behaviour when no longer your partner - will be covered in Section 2: VIOLENCE BY PEOPLE YOU KNOW WELL

- (•) Yes
- ( ) No
- ( ) Don't know / Can't remember
- ( ) Don't wish to answer

⇦ Back | Continue ⇨ | Reset these answers


New Zealand Crime & Safety Survey 2014
MINISTRY OF JUSTICE
Q167.420 How many times (since 1st January 2013)?
[Question was: Since 1st January 2013, has any partner actually used force or violence on you in some way, or deliberately hit you with something?]
PLEASE TYPE IN BELOW.
Range 1-999
Number 2
⇦ Back
Continue ⇨
Reset these answers


New Zealand Crime & Safety Survey 2014
MINISTRY OF JUSTICE
SECTION 2: VIOLENCE BY PEOPLE YOU KNOW WELL
We are also interested in learning more about people's experience of violence by people they know well.
International research suggests that this is much more common than violence by strangers. We need to know if this is so in New Zealand too.
This section focuses on violence by people you know well, including:
- Relatives
- Friends
- Family members
- Ex partners - husband / wife or partner with whom you were not in a marital or similar partnership at the time the violence happened
- Other people you know well.
⇦ Back
Continue ⇨



New Zealand **Crime & Safety Survey** 2014

MINISTRY OF JUSTICE Tāhū o te Ture

CBG PUBLIC SECTOR SURVEYING

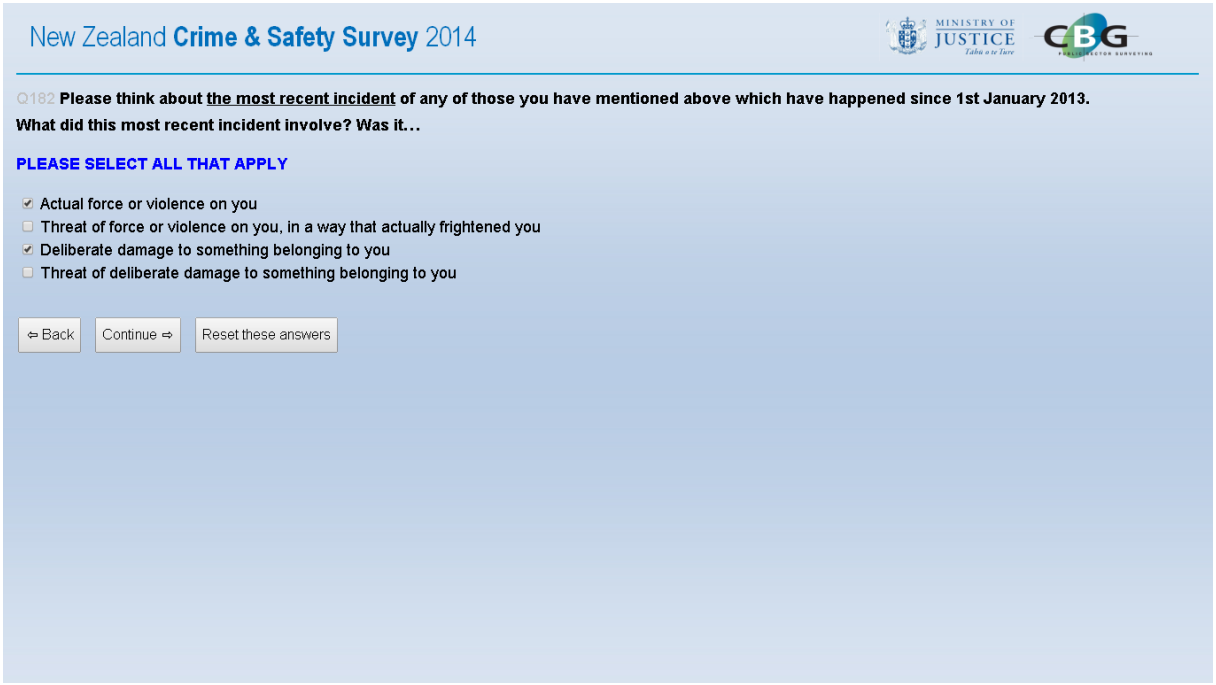
SECTION 3: SEXUAL INCIDENTS

Other stressful events are unwanted sexual advances. These are not always reported to the police, or even discussed with family, partners or friends. The person making these advances is not always a stranger, but can be a partner, friend or family member. Such experiences can occur any time in one's life - even as a child.

This next section asks about some unwanted sexual experiences you may have had. You may find the questions disturbing or distressing. But it is important to ask them because it will give a clearer idea of how often such incidents happen. Remember the information you give is confidential. Even if you have not experienced any incidents of the type we are going to ask about, we still need to know that. In this case, you will only need to answer a few questions.

We start with the most serious type of incident.

⇦ Back Continue ⇨



New Zealand **Crime & Safety Survey** 2014

MINISTRY OF JUSTICE Tāhū o te Ture

CBG PUBLIC SECTOR SURVEYING

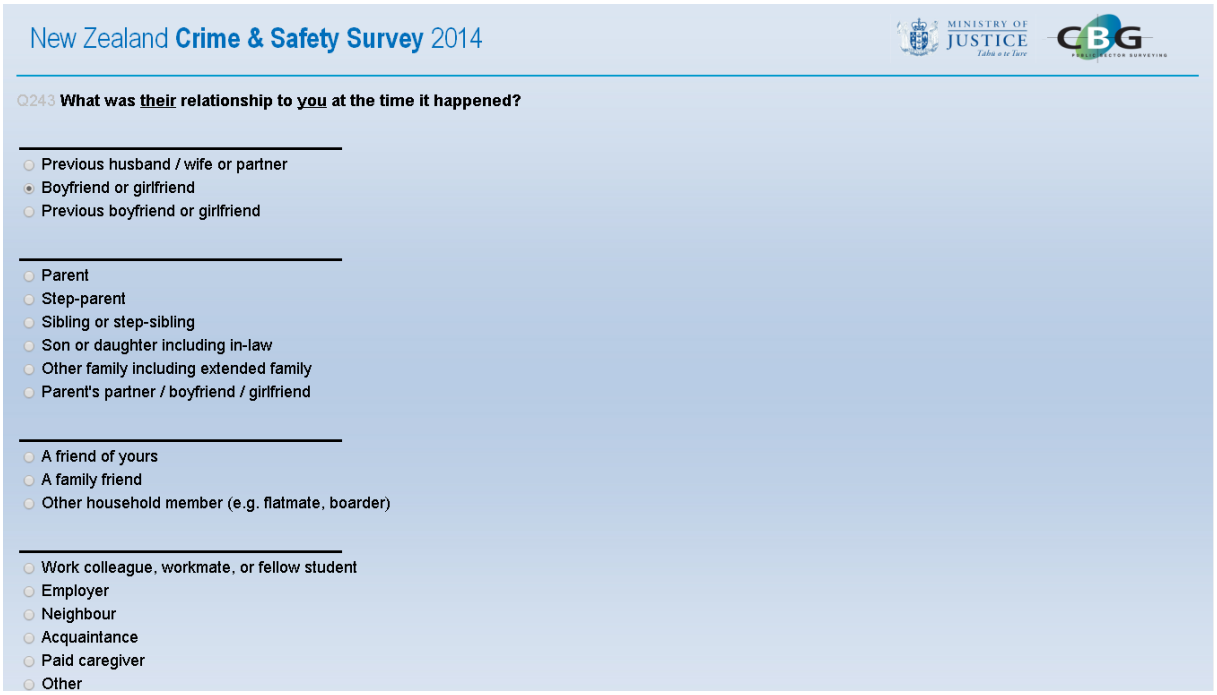
Q182 **Please think about the most recent incident of any of those you have mentioned above which have happened since 1st January 2013.**

What did this most recent incident involve? Was it…

PLEASE SELECT ALL THAT APPLY

- [x] Actual force or violence on you
- [ ] Threat of force or violence on you, in a way that actually frightened you
- [x] Deliberate damage to something belonging to you
- [ ] Threat of deliberate damage to something belonging to you

⇦ Back Continue ⇨ Reset these answers



New Zealand **Crime & Safety Survey** 2014

MINISTRY OF JUSTICE

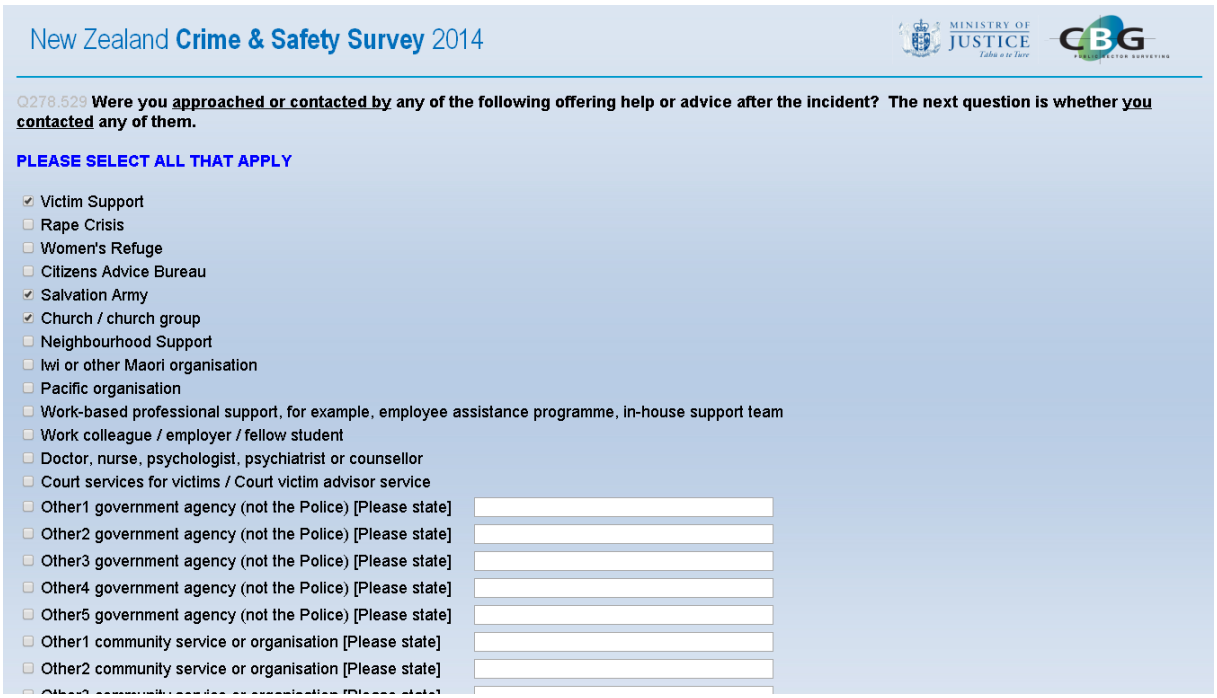
Q243 **What was their relationship to you at the time it happened?**

- Previous husband / wife or partner
- Boyfriend or girlfriend
- Previous boyfriend or girlfriend

- Parent
- Step-parent
- Sibling or step-sibling
- Son or daughter including in-law
- Other family including extended family
- Parent's partner / boyfriend / girlfriend

- A friend of yours
- A family friend
- Other household member (e.g. flatmate, boarder)

- Work colleague, workmate, or fellow student
- Employer
- Neighbour
- Acquaintance
- Paid caregiver
- Other



New Zealand **Crime & Safety Survey** 2014

MINISTRY OF JUSTICE

Q278.529 **Were you approached or contacted by any of the following offering help or advice after the incident? The next question is whether you contacted any of them.**

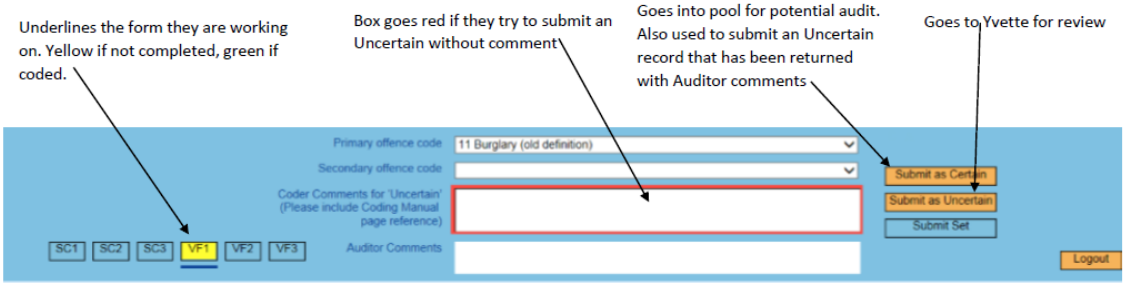
PLEASE SELECT ALL THAT APPLY

- [x] Victim Support
- [ ] Rape Crisis
- [ ] Women's Refuge
- [ ] Citizens Advice Bureau
- [x] Salvation Army
- [x] Church / church group
- [ ] Neighbourhood Support
- [ ] Iwi or other Maori organisation
- [ ] Pacific organisation
- [ ] Work-based professional support, for example, employee assistance programme, in-house support team
- [ ] Work colleague / employer / fellow student
- [ ] Doctor, nurse, psychologist, psychiatrist or counsellor
- [ ] Court services for victims / Court victim advisor service
- [ ] Other1 government agency (not the Police) [Please state]
- [ ] Other2 government agency (not the Police) [Please state]
- [ ] Other3 government agency (not the Police) [Please state]
- [ ] Other4 government agency (not the Police) [Please state]
- [ ] Other5 government agency (not the Police) [Please state]
- [ ] Other1 community service or organisation [Please state]
- [ ] Other2 community service or organisation [Please state]

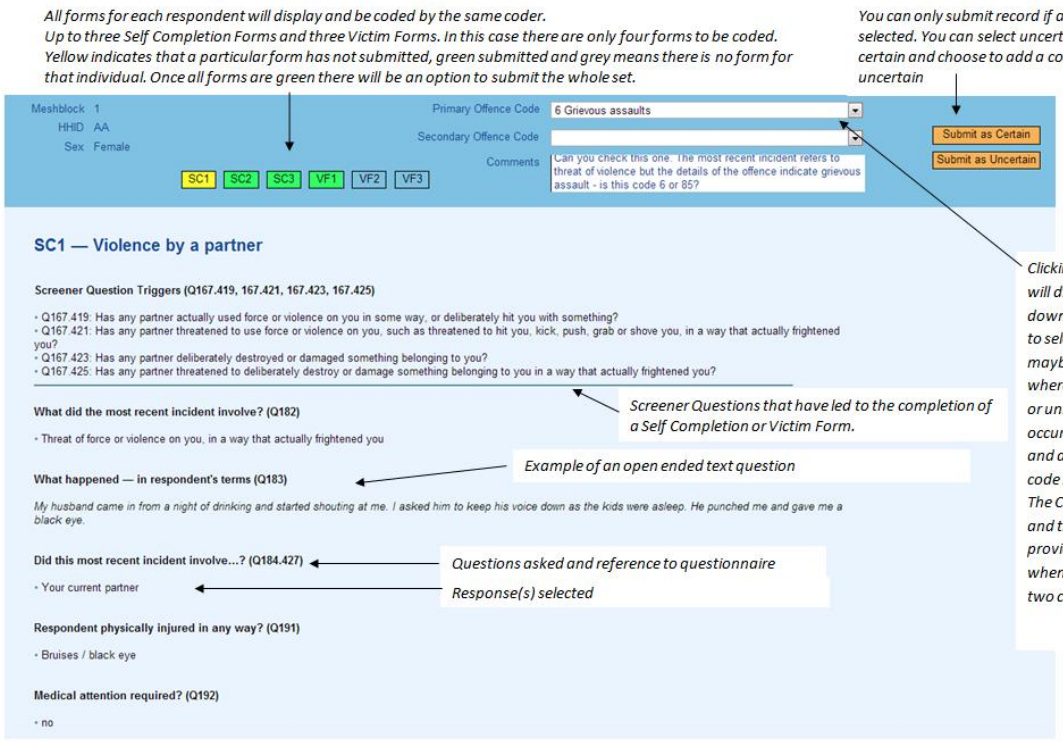

New Zealand Crime & Safety Survey 2014
MINISTRY OF JUSTICE
Tāhū o te Ture
CBG
Q409 Record how much you helped the respondent with the self completion section.
Self completion done by respondent
Respondent required help with one or two questions
Respondent required help with more than one or two questions but less than half
Respondent required help with more than half the questions but not all
Respondent required help with all or nearly all of the questions
⇦ Back
Continue ⇨
Reset these answers

Appendix D: Coding interface

CODER INTERFACE



Label	n	
Pool	281	Total number in pool ready for initial coding
My Uncertain Returned	0	Number of outstanding Uncertain/failed audits waiting for review, for individual Coder
My Week	28	Number completed that week
My Submitted	28	Total number of certain submitted by individual Coder overall
My Uncertain	17	Total number of Uncertain submitted by individual Coder overall



CBG Public Sector Surveying

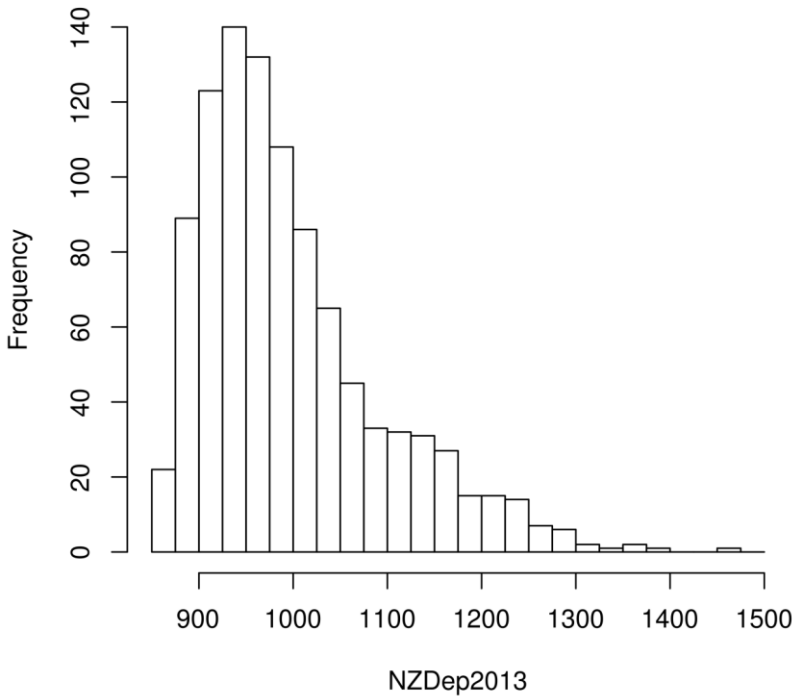
Appendix E: Survey weights

Non-response predictor variables

This section provides more detailed information about 2 of the predictor variables used in the non-response model, namely the deprivation index and transformed crime rate. NZDep2013 is an area-level summary of deprivation measures. It is widely used in New Zealand social research (Salmond and Crampton 2012), and is described fully by Atkinson, Salmond and Crampton (2014). It is defined at meshblock level, so everyone in the same meshblock has the same deprivation score, and is a right-skewed continuous variable with a national mean of 1000 and standard deviation of 100.

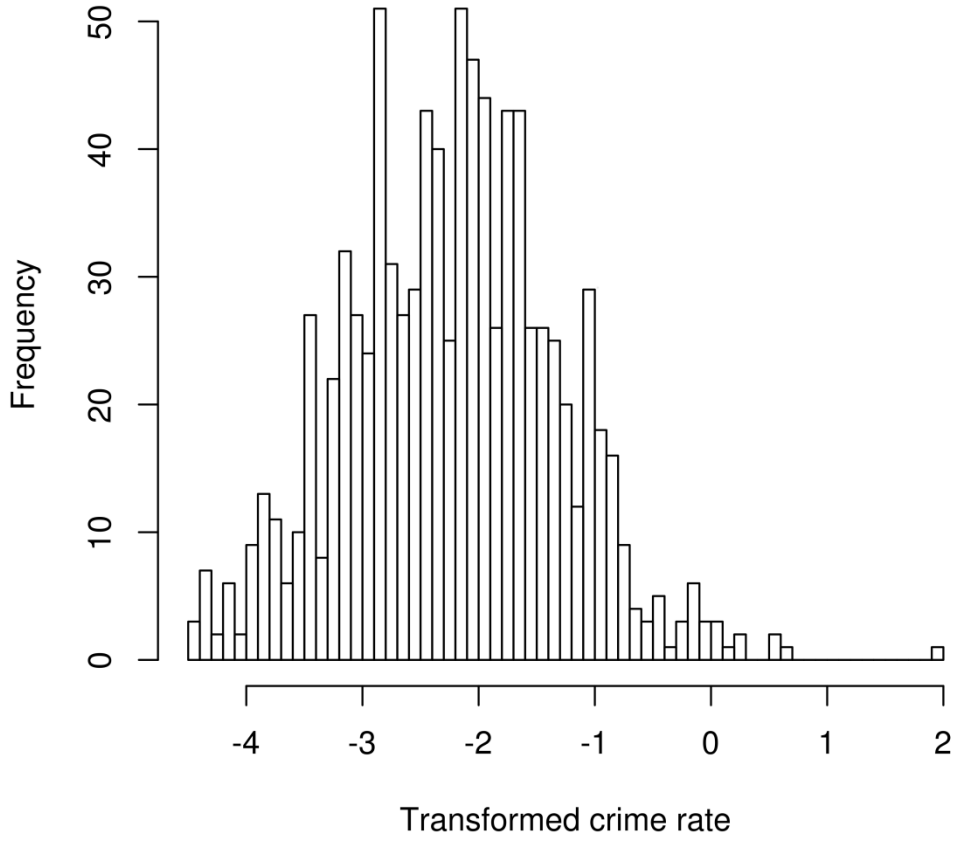
Figure E1 shows a histogram of its distribution over the 1000 meshblocks selected for the 2014 NZCASS, where it ranges from 853 to 1454.

Figure E1: Values of the deprivation index NZDep2013 for the meshblocks (PSUs) selected for the NZCASS



The transformed crime rate is intended as a broad indicator of the level of crime recorded in each area. The crime rate was derived by dividing the number of offences recorded by Police in that area by the population resident in that area. This was done both for Police station areas (as in 2009) and for meshblocks. (Meshblock-level data was not yet available in 2009.) Both sets of crime rates were strongly skewed to the right. The logarithm of the crime rates was taken and truncated for use in the non-response model. Both crime rates had non-significant p-values and were dropped from the model during stepwise selection. Figure E2 shows a histogram of the transformed and truncated crime rate for the 1000 meshblocks in the 2014 NZCASS sample.

Figure E2: Transformed crime rate



Sample and population profiles

Table E1 compares the 2014 NZCASS sample profile, unweighted and after each stage of the weighting process, with the corresponding population proportions across weighting control variables.

Table E1: Sample and population profiles for variables used in post-stratification and raking, by year

2014						
	Unweighted sample	Initial probability weights only	After non-response adjustment	Final weights (after post-stratification or raking)	Population (2014 estimates)	Difference between initial & final weights
Urbanisation						
Auckland	24.7%	26.7%	28.7%	27.7%	27.7%	1.0
Other metropolitan cities	17.9%	20.3%	20.2%	21.3%	21.3%	1.0
Other main urban areas	25.3%	22.5%	21.6%	21.0%	21.0%	1.5
Secondary urban areas	18.1%	16.3%	15.9%	15.3%	15.3%	1.0
Rural/minor urban areas	13.9%	14.3%	13.6%	14.8%	14.8%	0.5
Age by gender						
Males 15–24	4.9%	6.8%	6.8%	9.1%	9.1%	2.3
Males 25–39	9.8%	10.3%	10.4%	11.3%	11.3%	1.0
Males 40–59	16.2%	16.5%	16.5%	16.3%	16.3%	0.2
Males 60–69	6.6%	6.7%	6.7%	6.3%	6.3%	0.4
Males 70+	6.5%	6.5%	6.5%	5.4%	5.4%	1.1
Females 15–24	6.1%	7.9%	8.0%	8.7%	8.7%	0.8
Females 25–39	13.2%	12.2%	12.2%	12.1%	12.1%	0.1
Females 40–59	19.4%	18.8%	18.8%	17.6%	17.6%	1.2
Females 60–69	8.4%	7.3%	7.3%	6.6%	6.6%	0.7
Females 70+	8.8%	6.9%	6.9%	6.6%	6.6%	0.3
Māori by age by gender						
Māori males 15–39	5.7%	2.0%	1.9%	3.4%	3.4%	1.4
Māori males 40–59	5.5%	1.5%	1.5%	2.0%	2.0%	0.5
Māori males 60+	2.9%	0.5%	0.6%	0.8%	0.8%	0.3
Māori females 15–39	8.6%	2.7%	2.6%	3.7%	3.7%	1.0
Māori females 40–59	7.4%	2.1%	2.0%	2.2%	2.2%	0.1
Māori females 60+	4.1%	0.9%	0.9%	0.9%	0.9%	0.0
Non-Māori	65.9%	90.2%	90.5%	87.0%	87.0%	3.2
Pacific						
Pacific	4.5%	6.4%	6.6%	6.3%	6.3%	0.1
Non-Pacific	95.5%	93.6%	93.4%	93.7%	93.7%	0.1
Asian						
Asian	6.7%	11.4%	11.9%	12.3%	12.3%	0.9
Non-Asian	93.3%	88.6%	88.1%	87.7%	87.7%	0.9
Average change between initial and final weights						**0.8**

2009						
	Unweighted sample	Initial probability weights only	After non-response adjustment	Final weights (after post-stratification or raking)	Population (2009 estimates)	Difference between initial & final weights
Urbanisation						
Auckland	27.0%	28.6%	28.6%	26.8%	26.8%	1.8
Other metropolitan cities	21.7%	22.8%	22.6%	21.9%	21.9%	0.9
Other main urban areas	22.0%	19.9%	20.2%	20.4%	20.4%	0.5
Secondary urban areas	15.8%	15.1%	15.2%	15.0%	15.0%	0.1
Rural/minor urban areas	13.6%	13.7%	13.4%	15.9%	15.9%	2.2
Age by gender						
Males 15–24	5.7%	7.4%	7.4%	9.4%	9.4%	2.0
Males 25–39	11.3%	10.9%	11.0%	12.1%	12.1%	1.2
Males 40–59	15.4%	16.2%	16.0%	16.6%	16.6%	0.4
Males 60–69	5.8%	5.6%	5.6%	5.6%	5.6%	0.0
Males 70+	5.1%	4.9%	4.9%	4.9%	4.9%	0.0
Females 15–24	6.4%	7.7%	7.8%	9.0%	9.0%	1.3
Females 25–39	15.4%	13.4%	13.5%	12.9%	12.9%	0.5
Females 40–59	20.3%	21.5%	21.5%	17.5%	17.5%	4.0
Females 60–69	7.2%	6.5%	6.4%	5.8%	5.8%	0.7
Females 70+	7.5%	5.9%	5.9%	6.3%	6.3%	0.4
Māori by age by gender						
Māori males 15–39	6.4%	2.4%	2.5%	3.5%	3.5%	1.1
Māori males 40–59	4.2%	1.4%	1.4%	1.8%	1.8%	0.4
Māori males 60+	1.8%	0.5%	0.5%	0.6%	0.6%	0.1
Māori females 15–39	9.1%	3.1%	3.2%	3.8%	3.8%	0.7
Māori females 40–59	6.1%	2.1%	2.2%	2.0%	2.0%	0.1
Māori females 60+	2.5%	0.7%	0.7%	0.7%	0.7%	0.0
Non-Māori	69.9%	89.7%	89.4%	87.5%	81.2%	2.2
Pacific						
Pacific	4.0%	5.8%	6.2%	6.1%	6.1%	0.3
Non-Pacific	96.0%	94.2%	93.8%	93.9%	93.9%	0.3
Asian						
Asian	7.4%	11.9%	11.9%	10.7%	10.7%	1.2
Non-Asian	92.6%	88.1%	88.1%	89.3%	89.3%	1.2
Average change between initial and final weights						**0.9**

2006						
	Unweighted sample	Initial probability weights only	After non-response adjustment	Final weights (after post-stratification or raking)	Population (2001 Census or 2006/estimates / projections)	Difference between initial & final weights
Urbanisation						
Auckland	22.4%	23.6%	26.2%	26.6%	26.6%	3.0
Other metropolitan cities	18.1%	19.2%	18.4%	21.6%	21.6%	2.4
Other main urban areas	24.2%	24.2%	24.8%	22.0%	22.0%	2.2
Secondary urban areas	7.6%	7.5%	6.9%	6.6%	6.6%	0.9
Rural/minor urban areas	27.8%	25.5%	23.7%	23.2%	23.2%	2.3
Age by gender						
Males 15–24	5.1%	7.3%	7.3%	9.5%	9.5%	2.2
Males 25–39	9.9%	9.8%	9.8%	12.5%	12.5%	2.7
Males 40–59	14.4%	14.8%	14.8%	16.8%	16.8%	2.0
Males 60–69	5.8%	5.5%	5.5%	5.1%	5.1%	0.4
Males 70+	5.5%	4.9%	4.9%	4.7%	4.7%	0.2
Females 15–24	7.8%	9.2%	9.4%	9.0%	9.0%	0.2
Females 25–39	16.8%	15.3%	15.4%	13.3%	13.3%	2.0
Females 40–59	19.8%	20.9%	20.8%	17.4%	17.4%	3.5
Females 60–69	6.6%	6.0%	5.9%	5.3%	5.3%	0.7
Females 70+	8.4%	6.2%	6.1%	6.3%	6.3%	0.1
Māori by age by gender						
Māori males 15–39	5.6%	4.2%	4.3%	3.8%	3.8%	0.4
Māori males 40–59	3.7%	2.4%	2.5%	1.8%	1.8%	0.6
Māori males 60+	2.0%	1.0%	1.1%	0.6%	0.6%	0.4
Māori females 15–39	11.6%	7.4%	7.7%	3.9%	3.9%	3.5
Māori females 40–59	6.1%	4.0%	4.1%	2.0%	2.0%	2.0
Māori females 60+	2.3%	1.3%	1.3%	0.6%	0.6%	0.7
Non-Māori	68.6%	79.7%	78.9%	87.3%	87.3%	7.6
Pacific						
Pacific	4.1%	5.4%	5.6%	5.7%	5.7%	0.3
Non-Pacific	95.9%	94.6%	94.4%	94.3%	94.3%	0.3
Asian						
Asian	5.1%	7.8%	8.2%	9.4%	9.4%	1.6
Non-Asian	94.9%	92.2%	91.8%	90.6%	90.6%	1.6
Average change between initial and final weights						**1.7**

*Difference expressed in absolute percentage points difference

As expected, the weighted sample profiles for these variables match the population figures precisely for all 3 NZCASS iterations.

Histograms of survey weights

Figure E3: Histogram of the initial household weights, untransformed

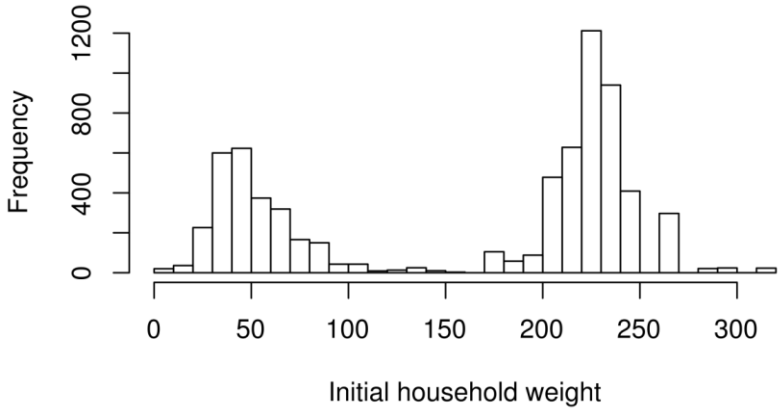
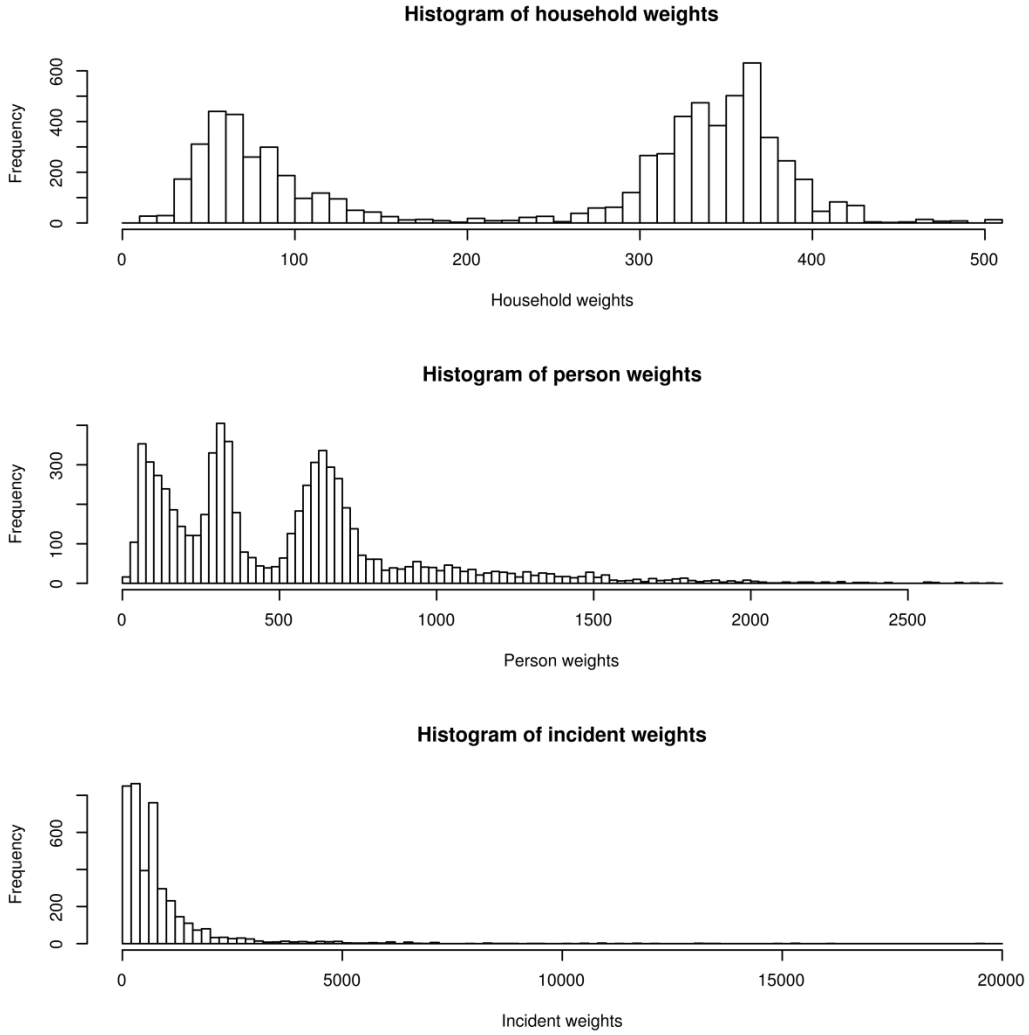


Figure E4: Histograms of the final untransformed household, person and incident weights



Design effects

The design effect for a statistic is the ratio of its variance estimate to its sampling variance calculated assuming that the data came from a simple random sample. It indicates how much effect the sample design, combined with the analysis techniques used (e.g. weighting and imputation), has had on the effective sample size. Design effects greater than one are common, and show that each interview provides less information than would an interview from a simple random sample.

Design effects were calculated for over 250 survey outputs, including incidence and prevalence estimates, across the last three iterations of NZCASS. In NZCASS 2014, the median design effects for analyses using person weights was 1.71, with a lower quartile of 1.52 and an upper quartile of 1.95. Design effects for the same outputs from NZCASS 2009 were generally smaller but had a similar interquartile range, whereas design effects for NZCASS 2006 statistics were usually even lower but were much more variable,[99] with the upper quartile of 2.08 being higher than those in 2009 and 2014. These increases in the design effects over time reflect changes to the sample design, primarily the decreased first-stage sampling fraction resulting from changing PSUs from Nielsen Area Units (in 2006) to meshblocks, and the booster sample being sourced from the electoral roll in 2014 instead of screening all dwellings. The following table summarises the distribution of design effects for each of the three surveys.

Table E2: Quartiles of design effects from NZCASS 2006, 2009 and 2014

	Minimum	Lower quartile	Median	Upper quartile	Maximum
2006	0.33	0.94	1.30	2.08	8.34
2009	0.49	1.37	1.53	1.75	2.80
2014	1.03	1.52	1.71	1.95	3.32

Variance estimates were calculated using a grouped jackknife in 2009 and 2014 and using Fay's method of balanced repeated replicates (BRR) in 2006. (BRR was used in NZCASS 2006 to handle the large first-stage sampling fraction and stratification.) The jackknife is known to generally give better variance estimates than BRR for the types of statistics produced in NZCASS (Judkins, 1990), although Fay's method can mitigate this. BRR can also suffer from instability when denominators are highly variable at the stratum level. If PSUs are grouped to form half-samples within strata, as in NZCASS 2006, BRR variance estimates may be biased for some variables if the half-samples are poorly balanced. The number of replicates used was increased from 28 in NZCASS 2006 to 100 in NZCASS 2009 and 2014. These issues explain the greater variability of NZCASS 2006 design effects.

Together with the smaller sample sizes achieved in NZCASS 2006, the greater variability of design effects in 2006 can yield substantially larger margins of error than in NZCASS 2009 and 2014 for a small proportion of outputs. For instance, the margin of error for the proportion of people who said the prison service is doing a good or excellent job in NZCASS 2006 was 3.3 percentage points (reflecting a design effect of 4.8), compared to 1.5 and 1.8 percentage points in 2009 and 2014 respectively (and corresponding design effects of 1.4 and 1.8).

[99] A more extensive investigation of design effects in NZCASS 2006 is described on pages 38-39 of Reilly and Sullivan (2008).

Appendix F: Imputation

The most recent incident and the estimation of interpersonal violence

Introduction

Within the CASI section, there are 3 parts which collect information about:

1. incidents by a current partner
2. incidents by a person well known
3. sexual incidents.

A victim form is completed for 1 incident in each of these 3 parts (unless no incidents are reported in a part). When the respondent reported experiencing more than 1 incident in a part, they were asked to complete the victim form thinking about the most recent incident.

The weighting and imputation methods used to handle the CASI victim form data are similar to those used for the CAPI victim forms. However, the selection of the most recent CASI incident in each part (in contrast to random selection of CAPI incidents) means that further assumptions are made when analysing CASI victim form data and the associated offence types.

Is there an undercount of intimate partner violence?

Ex-partners are included within the 'people well known' part of the CASI section along with a range of other relationships types such as other family members, friends, neighbours and workmates.

Where a respondent has experienced a number of incidents, they complete a victim form for the most recent incident. Because the most recent incident may not have been committed by an ex-partner (part of the intimate partner grouping), there is a question by some users about whether violence by ex-partners (and hence intimate partners) is undercounted. This same question could also be posed for the other relationship types.

This is in contrast to the first part of the CASI section which only gathers data on incidents committed by current partners. As such, we have details about an incident committed by a current partner from all respondents who experienced incidents. Despite incidents committed by current partners facing no competition for selection from incidents by other types of offenders, they do face competition from other incidents committed by current partners. If this was not adjusted for, these incidents could also be subject to some undercount.

The selection process is essentially the same in each CASI part, and the same statistical methods for handling this apply equally to incidents from all sections, as described below.

Assessing any potential undercount

In summary, the following analysis shows that weighting and imputation procedures count ex-partner violence correctly.

Weighting

The data collected in CASI victim forms does not represent a random sample of all incidents experienced since only information on the most recent incident is collected. It is instead skewed towards the types of incidents that tended to be experienced by a respondent only once or a few times during the recall period, and away from those that tend to be experienced together with many other incidents in that section.

As a result, unweighted analysis of CASI victim form data would often be subject to substantial bias. The reported analyses of CASI victim form data are always weighted using the incident weights described in Chapter 10, and these incorporate the number of incidents reported in the screener questions in each part as a weighting factor.

Selecting the most recent incident is a quasirandom selection process. It is intended to both be easy for the respondent to follow (by just thinking about the most recent incident) and provide a reasonable approximation to random selection for many variables. However it would not be close to random for variables that have a strong statistical relationship with the date of incident. For example if for some reason the respondent was more likely to report an incident experienced in 2014 to the Police, compared to an incident experienced in 2013, then the data on reporting to Police would not be a random selection of incidents.

The interpersonal violence relationship framework uses information on the respondent's knowledge of or relationship to the offender(s). Table F1 demonstrates the reporting of this information is unrelated to the year the incident occurred. There are some minor differences in percentages, but these are not statistically significant. This table shows no evidence that selection of the most recent incident, rather than randomly selecting an incident, would cause bias for analyses using offender relationship.

Table F1: Unweighted offender relationship profile from CASI victim forms, by the year each incident occurred

Relationship to offender	2013	2014
Family	63%	67%
Intimate partner	43%	44%
Current partner	32%	36%
Ex-partner	11%	8%
Family excluding intimate partners	22%	23%
Not family	38%	33%
People known excluding family	34%	29%
Strangers	5%	4%

Since there is no evidence of bias in selecting the most recent incident for the respondent's knowledge of or relationship to the offender(s), it is a reasonable assumption that CASI incidents are selected at random and are weighted by the number of incidents reported at the screener questions in that part. This weighting of incidents allows for the competition of incidents within the 'people well known' part, and the relationship profile is not damaged by this competition.

Imputation

Imputation methods can similarly adjust for random incident selection, for instance by including incident selection probabilities in the imputation model. Offender relationship is imputed (along with the offence codes) conditioning on the screener question the incident was reported at.

There are differences in imputation for CAPI and CASI. For CAPI, we know the screener question that the victim form incident relates to. Whereas for CASI, since the incident is the most recent, we do not know which screener question the incident relates to. This means for CAPI, the imputation process is simpler and the incident selection probabilities can be incorporated directly. For CASI there is an extra step required to impute the screener question the incident came from. Once this extra step is completed, the imputation process reflects the incident selection probabilities for both CAPI and CASI, and the relationship profile after imputation is similar with the weighted relationship profiles before imputation. In technical language, this means the design is ignorable (Gelman *et a*l, 2004).

Although the weighting and imputation processes incorporate incident selection in different ways, both allow for the selection of one incident from each CASI part, and the relationship profiles resulting from each method were expected to be similar. Table F2 compares the unweighted offender relationship profile of all NZCASS 2014 CASI incidents after imputation with the corresponding weighted and unweighted profiles for CASI victim forms before imputation. (The weights used here are simply the number of incidents recorded in that CASI part's screener questions.)

Table F2: Comparisons of unweighted, weighted and imputed offender relationship profiles from CASI incidents

Relationship to offender	CASI victim forms (before imputation)		All CASI incidents (after imputation)
	Unweighted	Weighted	Imputed
Family	65%	85%	80%
Intimate partner	44%	67%	61%
Current partner	34%	57%	54%
Ex-partner	10%	9%	7%
Family excluding intimate partners	22%	22%	19%
Not family	34%	19%	19%
People known excluding family	31%	19%	18%
Strangers	4%	0%	1%

The imputed offender relationship profile is similar to the weighted profile, while both differ markedly from the unweighted profile. In percentage point difference terms, the difference between 9% and 7% for the ex-partner row is small, and some minor differences are expect due to steps undertaken in imputation.

This confirms that the imputation of offender relationship has correctly adjusted for the selection of one incident from each part of the CASI section. In particular, the imputation of relationship types gathered in the second CASI part seems just as effective as the imputation of incidents by current partners.

Details of relevance imputation models

Tables F3 and F4 provide details of the Bayesian logistic regression models used for imputing relevance for CAPI and CASI incidents in the 2014 NZCASS. Relatively few of the parameter estimates are statistically significantly different from zero. For the CAPI model, such parameters include gender, home ownership, landlord, never having been married, and several screener questions. For the CASI model, they include just two screener questions (distressing sexual touching, and threats of damage from non-partners).

The models were fit using the bayesglm function in the arm R package (Gelman *et al* 2009), using the default prior distribution for all parameters (namely the Cauchy distribution with centre 0 and scale 2.5, except for the intercept which had a prior scale of 10). This fits the model using an approximate expectation–maximisation algorithm.

Table F3: Relevance imputation model for CAPI incidents

Parameter	Parameter estimate	Standard error
Intercept	−1.8354	0.5786
Screener – Theft from a vehicle	−0.0923	0.2516
Screener – Damage to a vehicle	0.5324	0.2327
Screener – Attempt to break into your home/garage	0.7903	0.2525
Screener – Unlawful entry into your home/garage	−0.2163	0.2707
Screener – Theft from outside property over $10	−0.1529	0.2441
Screener – Theft from inside your home by someone allowed to be there	−0.0877	0.2758
Screener – Deliberate damage to property belonging to your household	−0.1247	0.2504
Screener – Assault on you	−0.0350	0.3083
Screener – Threat of assault on you	0.9578	0.2399
Screener – Other damage to your personal property	1.0134	0.3459
Screener – Threat to damage your personal property	1.2821	0.3687
Screener – Theft or attempted theft of something you were carrying	−0.1799	0.3897
Screener – Theft of your personal property	1.1052	0.2571
Screener – Other offences	1.7148	0.2545
Household composition – 1 parent with children	0.0089	0.2028
Household composition – Couple without children	0.0748	0.1792
Household composition – Couple with children	−0.0795	0.2373
Household composition – Family, not elsewhere classified	0.2319	0.2410
Household composition – Flatmates	0.1602	0.2595
Household composition – Other	0.5323	0.3242
Gender – Female	0.2380	0.0861
Age – 25–39	0.0271	0.1601
Age – 40–59	0.0671	0.1697

Parameter	Parameter estimate	Standard error
Age – 60+	0.0220	0.2225
Marital status – Never been married or in a civil union	−0.2651	0.1298
Marital status – Widow/widower	0.0967	0.2261
Marital status – Divorced/separated	−0.0715	0.1544
Urbanisation – Metropolitan cities except Auckland	0.0670	0.1204
Urbanisation – Other main urban areas	−0.0414	0.1142
Urbanisation – Other urban areas	0.1985	0.1263
Urbanisation – Rural areas	0.1746	0.1534
NZDep13 score	0.0004	0.0004
Employment status – Home or caring duties	0.1546	0.1584
Employment status – Retired	−0.1880	0.1925
Employment status – Out of work or unable to work	0.1406	0.1420
Employment status – Studying	0.0681	0.1824
Ethnicity – European	−0.0521	0.1191
Ethnicity – Māori	−0.1953	0.1069
Ethnicity – Other	0.0602	0.1553
Household size	0.0245	0.0766
Tenure and landlord – Rented, not from a private landlord	−0.2516	0.1000
Tenure and landlord – Owned	−0.4903	0.1702
Tenure and landlord – Other/Don’t know/Refused	0.4793	0.4808

Table F4: Relevance imputation model for CASI incidents

Parameter	Parameter estimate	Standard error
Intercept	0.3149	1.4753
Screener – SC1 – Threat of force or violence	−0.6620	0.5569
Screener – SC1 – Deliberate damage	−0.2285	0.6328
Screener – SC1 – Threat of damage	−0.4534	0.6742
Screener – SC2 – Force or violence	0.5794	0.4164
Screener – SC2 – Threat of force or violence	0.2095	0.4665
Screener – SC2 – Deliberate damage	0.0196	0.6022
Screener – SC2 – Threat of damage	1.1890	0.5745
Screener – SC3 – Forced sexual intercourse	0.2404	0.8061
Screener – SC3 – Attempted forced sexual intercourse	−2.3107	1.4985
Screener – SC3 – Distressing sexual touching	−3.6815	1.6117
Screener – SC3 – Other sexual violence or threats	−2.0028	1.5122
pSC_1	−0.4939	0.6074
pSC_2	0.3557	0.6112
pSC_3	−0.2408	0.7042
Gender – Female	−0.0324	0.2432
Age – 25–39	−0.1861	0.3330
Age – 40–59	−0.5211	0.3831
Age – 60+	−0.7181	0.5985
Marital status – Never been married or in a civil union	−0.3615	0.3244
Marital status – Widow/widower	0.5426	0.7756
Marital status – Divorced/separated	−0.1155	0.3497
Urbanisation – Metropolitan cities except Auckland	0.3161	0.3589
Urbanisation – Other main urban areas	0.0567	0.3066
Urbanisation – Other urban areas	−0.1920	0.3483
Urbanisation – Rural areas	−0.2934	0.4175
NZDep13 score	−0.0011	0.0012
Employment status – Home or caring duties	−0.1157	0.3453
Employment status – Retired	0.2733	0.6635
Employment status – Out of work or unable to work	−0.2230	0.3260
Employment status – Studying	−0.3558	0.4075
Ethnicity – European	−0.2682	0.2725
Ethnicity – Māori	0.3817	0.2793
Ethnicity – Other	0.3409	0.4014

Parameter	Parameter estimate	Standard error
Household size	−0.0237	0.1327
Tenure and landlord – Rented, not from a private landlord	−0.0930	0.2519
Tenure and landlord – Owned	0.0867	0.3768
Tenure and landlord – Other/Don't know/Refused	0.8926	1.1357

Rate of missing information

Imputation had a fairly modest effect of the variances of overall victimisation rates, but had a more substantial effect on the variances of certain offence types, such as assaults. A standard diagnostic measure for multiple imputation is the 'rate of missing information' which shows how strongly the quantity being estimated is influenced by missing data.

Table F5 shows the percentage of missing information for the incidence and prevalence of selected offence types.[100] This indicates how strongly missing data affects the statistic being estimated, and is directly related to the efficiency of the estimate.

It is calculated as

$$\gamma = \frac{r+2}{(r+1)(v+3)}$$

where

$$v = 99\left(1 + \frac{\frac{1}{100}\sum_{j=1}^{100} \hat{v}_j(\hat{\theta})}{\left(1+\frac{1}{100}\right) var[\hat{\theta}_j]}\right)^2$$

is the degrees of freedom for the multiple imputation variance estimate, and

$$r = \left(1+\frac{1}{100}\right)\frac{\text{var}[\hat{\theta}]}{\frac{1}{100}\sum \hat{v}_{\text{JK},j}(\hat{\theta})}$$

is the relative increase in variance due to missing data.

[100] The figures for 2006 and 2009 differ from those published in previous Technical Reports. This is due to the changes made to the imputation process, such as the introduction of the Bayesian regression model and the increased number of imputations.

Table F5: Percentage of missing information for selected offence types

NZCASS iteration	2014		2009		2006	
	% of missing information for		% of missing information for		% of missing information for	
Offence type	Incidence	Prevalence	Incidence	Prevalence	Incidence	Prevalence
Assaults	10%	24%	12%	16%	11%	16%
Threats	11%	19%	19%	19%	14%	20%
Burglary	8%	7%	14%	8%	12%	9%
Vehicle offences	11%	12%	15%	12%	12%	10%
Sexual offences	6%	10%	10%	6%	7%	13%
Interpersonal violence	5%	14%	7%	10%	5%	10%
Personal offences	5%	13%	6%	10%	5%	10%
Household offences	6%	5%	10%	4%	8%	5%
All offences	5%	5%	7%	4%	6%	3%

Like all statistical techniques, multiple imputation relies on certain assumptions, including that proper imputation methods and models are used. A primary requirement is that the imputation process incorporates appropriate levels of variation in the imputed values, reflecting not just the observed distribution of the variable being imputed, but also uncertainty in the model parameters. It is also assumed that the imputation and analysis models are congenial (Meng 1994), ie that the analysis model can be derived from the imputation model. Uncongeniality could result in biased variance estimates, although variances are usually only slightly overstated if the analysis model is less complex that the imputation model.

Appendix G: NZCASS-ANZSOC concordance

Published comparable offences	ANZSOC concordance	Adjustment applied for NZCASS 2014
Thefts of vehicles	SCENE = TOTAL SCENE **> Theft and related offences > Motor vehicle theft and related offences** 1. Theft of a motor vehicle **> Theft and related offences > Motor vehicle theft and related offences > Illegal use of a motor vehicle** 2. unlawful takes motor vehicles (motor cars/trucks etc) 3. unlawful takes motor cycle (power cycles/scooters etc) 4. unlawful converts motor vehicles (motor cars/trucks etc) 5. unlawful converts motor cycle (power cycles/scooters etc) 6. attempted unlawful taking of a motor vehicle 7. other unlawful taking/conversion	Adjustment applied to offence level data to remove commercial vehicles. Exact statistics on this are not possible hence the proxy was applied to reduce the offence level data by 4.5%. This adjustment was obtained from the ratio of organisations to persons for this category in 2013 Police victim level data.
Thefts from vehicles/ vehicle interference	SCENE = TOTAL SCENE **> Theft and related offences > Motor vehicle theft and related offences > Illegal use of a motor vehicle** 1. unlawful gets into/upon motor vehicle/motor cycle 2. other unlawful interference/getting into **> Theft and related offences > Motor vehicle theft and related offences** 3. theft of motor vehicle parts or contents **> Property damage and environmental pollution > Property damage > Property damage, nec** 4. unlawful interferes motor vehicles (motor cars/trucks etc) 5. unlawful interferes motor cycle (power cycles/scooters etc)	Adjustment applied to offence level data to remove commercial vehicles. Exact statistics on this are not possible hence the proxy was applied to reduce the offence level data by 3.9%. This adjustment was obtained from the ratio of organisations to persons for this category in 2013 Police victim level data.
Burglary	SCENE = DWELLING 1. Unlawful entry with intent/burglary, break and enter **> Public order offences > Disorderly conduct > Criminal intent** 2. Armed with intent to commit burglary (firearm) 3. Armed with intent to commit burglary (other weapon) 4. Possession instrument for burglary (SOF) 5. Possess instruments for burglary 6. Disguised for burglary	No adjustment applied

Published comparable offences	ANZSOC concordance	Adjustment applied for NZCASS 2014
Robbery/thefts from the person	SCENE = TOTAL SCENE **> Robbery, extortion and related offences > Robbery > Aggravated robbery** 1. Aggravated robbery (cause GBH) firearm 2. Aggravated robbery (cause GBH) other weapon 3. Aggravated robbery (cause GBH) manual 4. Aggravated robbery (firearm) 5. Aggravated robbery (other weapon) 6. Aggravated robbery (manually) 7. Aggravated robbery (cause GBH) stabbing/cutting weapon 8. Aggravated robbery (stabbing/cutting weapon) 9. Other aggravated robbery 10. Robbery (by assault) 11. Assaults with intent to rob (firearm) 12. Assaults with intent to rob (other weapon) 13. Assaults with intent to rob (manually) 14. Assault intent to rob (stabbing/cutting weapon) 15. Assaults with intent to rob (with another person) 16. Assaults with intent to rob (cause GBH) 17. Other assaults with intent to rob 18. Compels execution of documents (firearm) 19. Compels execution of documents (other weapon) 20. Compels execution of documents (manually) 21. Compels execution of documents (by threats) 22. Compels execution of doc (stabbing/cutting weapon) 23. Other compelling execution of documents 24. Aggravated robbery (together with another person/s) **> Robbery, extortion and related offences > Robbery** 25. Non-aggravated robbery **> Theft and related offences > Theft (except motor vehicles)** 26. Theft from a person (excluding by force)	Victim level data used only where victim is aged 14 years and above to align to NZCASS responding population (15 years and above is the scope for NZCASS but as questions relate to the year prior, the respondent could be 14 at time of incident). Victim level data is more appropriate than offence level data to align to the collection of NZCASS personal offences. Victim level data is larger than offence level as there can be multiple victims per offence.
Assaults	SCENE = TOTAL SCENE 1. Acts intended to cause injury	Victim level data used only where victim is aged 14 years and above to align to NZCASS responding population. Note 15 years and above is the scope for NZCASS but as questions relate to the year prior, the respondent could be 14 years at time of incident. Victim level data is more appropriate than offence level data to align to the collection of NZCASS personal offences.

Appendix H: Historical investigations

In each iteration of the NZCASS and it's predecessor (the NZ National survey of crime victims 'NZNSCV') different things have been investigated for different reasons and as part of the project's due diligence. Because many things either do not change notably over time, or are only relevant to particular year of the survey, some of these investigations have only been conducted once. The following table summarises these investigations and where users can find more information.

What	Description	Year	Reference
Approximate effect of imputation of year of offence	Estimate of random error in incidence and prevalence rates due to date imputation	1996	Final Report to Victimisation Project Committee, Addendum to Appendix 2, pp 151-152.
Sensitivity analysis of imputation assumptions for the 1996 NZNSCV	Investigation of how sensitive 1995 victimisation estimates were to failures of various assumptions.	2001	Technical Report on 2001 NZNSCV, Chapter 7.
Response rate by interviewers experience	Analysis of fieldwork outcomes and response rates by interviewer experience for the main sample.	2006	2006 Technical Report, Appendix A1, Table A1.1
Sample and population profiles for other household characteristics	Percentages for unweighted sample, probability weights, after non-response adjustment, final weights after post stratification / ranking, population (census or estimates)	2006	2006 Technical Report, Appendix A2, Table A2.2
Sample and population profiles for other personal variables	Percentages for unweighted sample, probability weights, after non-response adjustment, final weights after post stratification / ranking, population (census or estimates)	2006	2006 Technical Report, Appendix A2, Table A2.3
Effect of area unit population changes	Analysis to assess levels of bias introduced through population changes between the 2001 census and the start of fieldwork in 2006	2006	2006 Technical report, Appendix A4.
Derivation of eligibility probability estimate	2006 household weights relied on an estimate of booster sample selection probabilities. Derivation of the formula for this estimate.	2006	2006 Technical report, Appendix A5.
Investigation of incident dates	Dates recorded in the victim forms show that the proportion of incidents reported in the reference year is lower than would have been expected if they were evenly spread (which is the main assumption made when imputing dates). Investigation to review possible causes and the effect on victimisation rates.	2006	2006 Technical report, Appendix A6.
Modelling issues and methodology	Potential pitfalls in multivariate analysis of victimisation risk. Description of modelling methodology used in NZCASS 2006.	2006	Understanding victimisation risk: findings from the NZCASS 2006 in an international context, Appendix B1.
Checks and analysis of Q44 responses in the CAPI form	Addition of 'refused' option to Q44 (can I confirm this incident happened in NZ and after 1 Jan 2008). Analysis of response options given question change.	2009	2009 Technical report, Appendix C.

What	Description	Year	Reference
Non-response predictor variables	Detailed information about the predictor variables used in the non-response model, namely the deprivation index and transformed crime rate.	2009	2009 Technical report, Appendix E, E1
Post-stratification and ranking	Explanation of the weighting techniques called post-stratification and raking, illustrating them using simple examples.	2009	2009 Technical report, Appendix E, E2
Sample and population profiles	Percentages for unweighted sample, probability weights, after non-response adjustment, final weights after post stratification / ranking, population (census or estimates)	2009	2009 Technical report, Appendix E, E3
Bias, variance and the heavy victimisation cut-off	Background information about how the cut-off value of 30 offences was chosen in 2006.	2009	2009 Technical report, Appendix F, F1

Terms and definitions

Adult(s)

Only people 15 years old and over were included in the NZCASS. The term 'Adult' refers to all NZCASS respondents.

Asian

1. This ethnicity category includes: Asian (not further defined), South-East Asian, Chinese, Indian and other Asian.
2. Also see ethnicity.

Assault

1. Where someone uses force against the respondent (including throwing objects), whether or not the respondent is injured. The respondent must have been the victim of the assault themselves for the offence to be counted in the NZCASS.
2. There were two types of assault:
    a. Grievous assault: Involves an aggravating element (where there was actual or intended injury or serious harm) in addition to the application of force.
    b. Other assault: any assault that is not grievous or indecent.

Burglary

1. When a person enters a building intending to commit a crime.
2. Burglary does not require forced entry. Includes thefts from enclosed spaces, such as yards. Does not cover theft by someone who had a right to be in the building.
3. The NZCASS only covers domestic burglary.
4. Burglary is a household offence in the NZCASS because the whole household, not just the respondent, were victims.

Buildings

1. Any 'building-type' structure on a property. This can include:
    a. homes or holiday homes
    b. flats, including 'common areas' (hallways, stairs, garages etc)
    c. caravans
    d. outhouses (connected)
    e. sleepouts
    f. garages, carports or boat sheds (on the property)
    g. toolsheds
    h. any 'enclosed yard' surrounding the dwelling.

CAPI (computer-assisted personal interviewing)

1. Where an interviewer enters answers to the survey directly into a laptop computer.
2. Used for the NZCASS in 2006, 2009 and 2014.

CASI (computer-assisted self-interviewing)

1. Where the survey participant puts their answers into a laptop computer themselves.
2. Ensures that survey responses remain private from the interviewer.

Classification

1. A way to group a set of related categories in a meaningful, systematic and standard format.
2. Term used to describe how demographic and geographic categories (rather than offences) are grouped together.
3. Where possible, NZCASS output aligns to Statistics NZ's classifications and standards.

Coercive control

For the NZCASS, this covers behaviours by a current partner that are intended to monitor, control, threaten or insult the victim.

Community violence

1. 'Community violence' is a term used by the World Health Organization (WHO). Within the context of NZCASS, 'Community violence' means violent offences by people who are not family.
2. Violent offences by people who are not family is a sub-set of the wider category 'interpersonal violence'.

Comparable offences/subset

The group of offences for comparisons between the NZCASS estimates and Police official crime statistics. The subset comprises of thefts of vehicles, thefts from vehicles / vehicle interference, burglary, robbery / theft from the person and assaults.

Concentration of crime

This measure shows how many times a person or household has experienced an offence. It is used to show multiple and repeat victimisation.

Confidence intervals

1. A statistical measure of an estimate's reliability (that is a measure of sampling error).
2. A confidence interval expresses the sampling error as a range of values in which the 'real' population value is estimated to lie.
3. The 95% confidence interval (CI) is used in NZCASS reporting, and is calculated as the estimate plus or minus the margin of error (MoE). This means CI = estimate ± MoE of the estimate.

Confrontational crimes

1. In the 2014 NZCASS, the term 'confrontational crimes' is replaced by 'interpersonal violence'. Confrontational crime was used in the 2006 and 2009 NZCASS.
2. Confrontational crimes include assaults, threats (to the person or property), abduction / kidnapping, robberies and sexual offences where the offence is committed by a partner or someone the victim knows well.
3. It excludes personal property damage.

Crime(s)

1. An action or omission which constitutes an offence and is punishable by law.
2. The words 'crime' and 'crimes' in the NZCASS is a general term which is used to describe a type of offence or group of offences. For example, 'vehicle crime' refers to a set of vehicle offences.

Current partner (violent offences by current partner)

1. 'Violent offences by current partners' (sometimes called 'current partner violence') refers to violence between individuals in a current intimate romantic relationship. They are a subset of 'interpersonal violence'.
2. A current partner can be a husband/wife, civil union partner, de facto partner, boyfriend girlfriend.
3. In the 2014 NZCASS, 'violent offences by current partners' covers a range of offences. See violent interpersonal offences.

Damage (vandalism)

Wilful damage to personal or household property:

a. includes (but is not limited to) damage, such as arson and graffiti
b. excludes incidents where they are a nuisance only (like letting down car tyres).

Dark figure of crime

1. Crime that happens in any year that is not reported in the NZCASS or to Police.
2. This crime is not recorded in the official Police statistics.

Dwellings

1. A 'permanent, private dwelling' can be:
    a. a separate house
    b. two or more separate houses
    c. flats joined together
    d. a flat or house joined to a business, shop, bach, crib or hut (as long as it's not attached to a work camp).
2. 'Permanent, private dwellings' include both occupied and unoccupied dwellings.
3. Temporary private dwellings – such as, caravans, cabins, tents or boats – were excluded from the survey.
4. All non-private dwellings – such as, hotels, motels, guest houses, boarding houses, hostels and motor camps – were excluded from the survey.

Estimates

The statistics produced by the NZCASS are called estimates because they are derived from a sample survey rather than from the entire New Zealand population.

Ethnicity

1. The ethnic group(s) the respondent said they identify with. People who identified with two or more ethnic groups were counted in each ethnic group they identified with.
2. Ethnicity is classified according to the 2-digit Ethnicity New Zealand Standard Classification (2005).

European

1. This ethnicity category includes: European (not further defined), NZ European, Other European and New Zealander/Kiwi.
2. Also see ethnicity.

Ex-partner (violent offences by ex-partners)

1. Violent offences by ex-partners (sometimes called 'ex-partner violence') refer to violence between people who are no longer in an intimate romantic relationship. This is a subset of 'interpersonal violence'.
2. Ex-partners can be ex-husbands, ex-wives, ex-civil union partners, ex-de facto partners, ex-boyfriends or ex-girlfriends.
3. Within the context of the 2014 NZCASS, 'violent offences by ex-partners' include a range of offences. See violent interpersonal offences.

Factors

1. Information collected by the NZCASS about the respondent (like where they live and how old they are) that helps us understand the relationship between different subgroups of people and a data item of interest (such as victimisation).
2. Factors are used to show if there are any differences between people in a particular group for a data item of interest, when compared to the New Zealand average (for example, whether the elderly are more or less likely to be victims of crime).
3. Also see risk.

Family excluding intimate partners (violent offences by family excluding intimate partners)

1. 'Violent offences by family excluding intimate partners' are violent offences committed by a family member who is not an intimate partner. They are a subset of 'interpersonal violence'.
2. These family members can be parents, step-parents, siblings, step-siblings, children, children-in-law, extended family or parents' partners.
3. In the 2014 NZCASS, 'violent offences by family excluding intimate partners' covers a range of offences. See violent interpersonal offences.

Family violence/violent offences by family

1. Violent offences by family (sometimes called 'family violence') refers to violence between people who are related. They are a subset of 'interpersonal violence'. Family can be:
    a. intimate partners
    b. parents or step-parents
    c. siblings or step-siblings
    d. children or children-in-law
    e. extended family
    f. parents' partners.
2. In the 2014 NZCASS, 'violent offences by family' cover a number of offences. See violent offences or the NZCASS data items list for more information.

General theft of personal property

See theft of personal property (general).

Grey figure of crime

1. Crime that is reported to the Police, but not counted in the official Police statistics.
2. Crime may be reported to Police but not recorded for a number of reasons, including:
    a. Police may not believe the victim
    b. there is not enough information to confirm whether a crime has occurred
    c. the victim may not want to take the matter further.

Groupings (offence groupings)

1. A way to group related offences together for analysis and reporting.
2. Different offence groupings were used for different reporting purposes. See Chapter 9 for further detail.

Household(s)

One or more people who usually live at the same private dwelling, who share common facilities and who define themselves as a household.

Household composition

1. A way to classify households based on the relationships of the people who usually live in the dwelling.
2. Households are defined by the number of families, the type of families and whether other related or unrelated people live there.
3. A family nucleus is a couple, with or without children, or one parent with their children, who all usually live together in the same household. The children do not have partners or children of their own living in the same household.
4. The term ‘other person(s)’ refers to people who are living in a household where there is at least one family nucleus, but the ‘other person(s)’ are not part of the family nucleus. The other person(s) may or may not be related to the family nucleus. For example, a ‘couple with child(ren) and other per son(s)’ household includes a family nucleus, consisting of a couple and their children, and at least one other person who is outside this family nucleus, such as a grandparent, uncle, aunt, flatmate, boarder, etc.
5. ‘Other multi-person household’ is a household with two or more people with no family nucleus. This includes households of related people who are not a family nucleus (such as siblings), households of unrelated people (such as flatmates) or a combination of related and unrelated people where there is no family nucleus.
6. For more information, see Statistics NZ’s standard classification of household composition.

Household offences/crimes

1. When the respondents' household is the victim of crime rather than the respondent personally, this is a household offence. In the NZCASS, household offences are:
    a. burglary
    b. theft – household property
    c. damage – household property
    d. theft – vehicle
    e. damage – vehicle
    f. thefts – from vehicle/vehicle interference
2. See the NZCASS data items list for more information.

Imputation

1. A statistical process done to fill in missing information.
2. Imputation is necessary because the NZCASS does not collect victim forms on all incidents reported in the NZCASS, but all the incidents still need to be counted in the overall incidence rates and prevalence rates.

Incidence rate

1. The average number of offences that happened in the reference year for each 100 households or adults.
2. It takes into account that some people and/or households are victimised more than once.
3. It does not give a good measure of someone's risk of victimisation because risk is not evenly distributed across the population – see prevalence rate.

Incident

1. An incident is something that happened, an event, or occurrence.
2. As part of the NZCASS, respondents are asked about different incidents that might have happened to them.
3. An incident may include one or more offences.
4. 'Incident' should not be confused with the 'incidence rate'.

Institutions

1. People living in institutions are not part of the NZCASS.
2. Institutions include hospitals, psychiatric institutions, prisons, barracks for the New Zealand armed forces and homes for the elderly.
3. Also see non-private dwellings.

Interpersonal violence/violent interpersonal offences

1. Interpersonal violence refers to violence between people. It covers:
    a. intimate partner violence (current and ex-partners)
    b. family violence
    c. violence by other known people (people who are not intimate partners or family)
    d. violence by strangers
2. In the 2014 NZCASS, 'violent interpersonal offences' include physical, sexual and threats and damage offences. See violent interpersonal offences.

Intimate partner violence/violent offences by intimate partners

1. Intimate partner violence is a subset of 'interpersonal violence'. It refers to violence between people in a current intimate romantic relationship or those who've been in an intimate relationship before. Intimate partners can be:
    a. current partners (husband, wife, civil union partner, de facto partner, boyfriend, or girlfriend)
    b. ex-partners (ex-husband, ex-wife, ex-civil union partner, ex-de facto partner, ex-boyfriend, or ex-girlfriend)
2. In the 2014 NZCASS, 'violent offences by intimate partners' includes a range of offences. See violent interpersonal offences.

Lifetime experience

Whether the offence has ever happened at some stage in their life.

Margin of error (MoE)

1. A statistical measure of an estimate's reliability (that is a measure of sampling error).
2. The 95% margin of error indicates there are about 19 chances in 20 that the value for the 'real' population will fall within the margin of error of the survey's estimate.
3. The 95% margin of error is used in NZCASS reporting, and is calculated as the t-value (approximately 1.96) multiplied by the standard error (MoE = t-value * standard error of estimate).
4. In the NZCASS, reporting estimates with an MoE between 10 and 20 percentage points are considered high and should be viewed with caution. Estimates with an MoE over 20 percentage points are suppressed as they are considered too unreliable for general use.
5. The margin of error is used to calculate confidence intervals.
6. The margin of error is used in NZCASS reporting as a sampling error measure for percentages.

Meshblock

1. A defined geographical unit varying in size from part of a city block to large areas of rural land.
2. Every part of New Zealand is classified to be within a certain meshblock.
3. The median number of people in each meshblock is 81 people in 2013.
4. Meshblocks are the smallest geographical unit defined by Statistics NZ to collect statistical data.
5. Can be combined to form larger statistical units, such as area units, territorial local authorities and regions.

Missing information

Missing information in the NZCASS has been filled in using the imputation process.

Multiple victimisation

1. When someone has experienced more than one offence of any type.
2. See also repeat victimisation.

Neighbourhood

1. In the NZCASS, the respondent defined this term themselves, unless they asked what it meant.
2. If the respondent asked what 'neighbourhood' meant, the interviewer said either:
   a. 'The streets around you' (NZCASS urban definition)
   b. 'Your district' (NZCASS rural definition).

Non-private dwellings

1. Those living in non-private dwellings are not part of the NZCASS.
2. This includes hotels, motels, guest houses, boarding houses, hostels, motor camps, staff quarters, hospitals and institutional complexes.
3. Also see institutions.

NZDep (NZ Index of Deprivation)

1. Standard measure of relative socio-economic deprivation in New Zealand.
2. It is derived from the Census using 9 variables to provide a deprivation score for each meshblock in New Zealand.
3. For further information refer to the University of Otago's NZDep information page.
4. In NZCASS, NZDep is converted into quintiles (5 evenly sized groups) where 1 represents the areas with the least deprived scores, and 5 the areas with the most deprived scores.

Offences/offence codes

1. An offence is a crime that meets a specific legal definition.
2. Offence codes are a number assigned to offences to help analysis and reporting for the NZCASS.
3. Where an offence is not measured by the NZCASS or the incident did not meet the legal definition of an offence, it was given the 'out of scope' code. For details, go to the NZCASS offence coding manual.

Offence type

An individual offence or group of similar offences used in reporting.

Other ethnic group

This ethnicity category includes: Middle Eastern, Latin American, African and 'other'. Also see ethnicity.

Out of scope

1. The types of people or offences not included as part of NZCASS.
2. The people excluded are children under 15 years and people who live in institutions or do not live in permanent private dwellings.
3. The offences excluded are 'victimless crimes' (such as drug offences), manslaughter and homicide, commercial crime, white-collar crime, crimes against businesses or public sector agencies, e-crime.

Pacific peoples

1. This ethnicity category includes: Pacific peoples (not further defined), Samoan, Cook Islands Māori, Tongan, Niuean, Tokelauan, Fijian and other Pacific peoples.
2. Also see ethnicity.

Partners

1. Includes current partners and ex-partners.
2. Used in regard to 'interpersonal violence'.

Partnered – legally registered

1. Where someone is legally married or is in a civil union at the time of interview.
2. See partnered – not legally registered.

Partnered – not legally registered

1. Where someone is in a relationship at the time of interview but not married or in a civil union.
2. See partnered – legally registered.

People known (excluding family) / violent offences by people known (excluding family)

1. Violent offences by people known (excluding family) are a sub-set of 'interpersonal violence'. The term refers to violence committed by people known to the victim, but who are not family members.
2. This includes friends, family friends, flatmates, colleagues, employers, neighbours, acquaintances and caregivers.
3. In the 2014 NZCASS, 'violent offences by people known (excluding family)' include a range of offences. See violent interpersonal offences.

People who are not family (violent offences by people who are not family)

1. Violent offences by people who are not family are a subset of 'interpersonal violence'. This is also called 'community violence'. The term refers to violence by a non-family member, including people known to the victim (but who are not family) and strangers.
2. In the 2014 NZCASS, 'violent offences by non-family members' covers a range of offences .See violent interpersonal offences.

Permanent, private dwellings

One of the following:

a. a separate house
b. two or more houses or flats joined together
c. a flat or house joined to a business or shop
d. a bach, crib or hut that are used as private dwellings (as long as it is not attached to a work camp).

Personal offences/crimes

1. When the respondent themselves was the victim of a crime (rather than their household), this is called a personal crime. In the NZCASS, personal offences are:
    a. sexual offences
    b. assault offences
    c. robbery
    d. threat – of force
    e. threat – to damage property
    f. damage – personal property
    g. theft – personal property
    h. theft – from the person.
2. See the NZCASS data items list for more information.

Violent (physical) offences

1. This category includes: assaults, abduction/kidnapping and robbery.
2. Comprises interpersonal violence offences, along with sexual offences and threats and damage offences.
3. See the NZCASS data items list for more information.

Prevalence rate

1. The prevalence rate shows the percentage of households or adults who were the victim of one or more offences in a given year.
2. Unlike incidence rates, prevalence rates do not take account of the number of times one person or household has been victimised.

Recall period

1. The time period that survey participants (respondents) are asked to take into account when they are answering a question.
2. The most frequently used recall period in the 2014 NZCASS is: ‘Since the 1st of January 2013’.
3. Also known as ‘reference period’.

Recording rate

1. Offences that were counted by the Police in the official statistics, compared to the number of offences victims said were reported to the Police (as recorded in the NZCASS).
2. Crime reported to the Police but not counted in the official statistics is known as ‘the grey figure of crime’.

Relative standard error (RSE)

1. A statistical measure of an estimate's reliability (that is a measure of sampling error).
2. The RSE is obtained by expressing the standard error as a percentage of the estimate, that is RSE = (standard error of the estimate / estimate) 100.
3. NZCASS reporting uses RSE reporting as a sampling error measure for count estimates and means.
4. In the NZCASS, reporting estimates with an RSE between 20% and 50% are considered high and should be viewed with caution. Estimates with an RSE over 50% are suppressed as they are considered too unreliable for general use.
5. The RSE is similar to but not to be confused with the relative sampling error used by Statistics NZ.

Repeat victimisation

1. When someone has experienced more than one offence of the same type.
2. See also multiple victimisation.

Reporting rate

1. The number of incidents that respondents say became known to the Police.
2. Incidents may have become known to the Police either because the victim or someone else reported it to the Police or because the Police themselves discovered the incident.

Risk (rates/factors)

1. The chances a person or household is likely to be a victim of crime.
2. Based on prevalence counts.
3. Many of the things that increase or decrease risk are interrelated and overlap.

Robbery

1. Where someone stole (or tried to steal) property from a respondent by using force/violence or threatening them with force/violence either during or immediately before a theft or attempted theft.
2. Violence may be used to extort the property stolen or to overcome resistance to it being stolen.
3. Aggravated robbery involves grievous bodily harm (serious injury to someone), two or more people working together in a robbery, or a robbery where the offender had a weapon.

Sampling error

1. Sampling error comes about because the NZCASS surveyed a group of people, not the entire New Zealand population.
2. As such, the results of the survey may differ from the results we would get if the entire New Zealand population had been included.
3. The size of the sampling error depends on the sample size, the size and nature of the estimate, and the design of the survey. The sampling error can be worked out mathematically.
4. There are a few related ways to express the difference between the survey estimate and the 'real' population:
    a. standard error
    b. margin of error (MoE),
    c. relative standard error (RSE)
    d. confidence intervals.
5. Sampling error is also taken into account in tests of statistical significance.

Sexual offences

1. Sexual offences include sexual violation, rape and indecent assault.
2. Information about sexual offences is collected in the third self-completion section of the survey, asked respondents about forced sexual intercourse, attempted forced sexual intercourse, distressing sexual touching, and other offences of sexual violence.
3. Forced sexual intercourse is defined as forced oral sex or forced anal or vaginal penetration.

Screening questions ('screeners')

Questions used to collect information:

a. on whether a survey respondent has experienced an incident or incidents that come within the scope of the survey
b. on how many times a particular incident has happened.

Social renters

1. Used in the 2006 and 2009 NZCASS to refer to people who rented from a local authority or Housing New Zealand, as well as people who rented but refused to say who they rented from, who gave an 'other' response, or who did not know their landlord.
2. Called 'rented – government (local and central)' in the 2014 NZCASS (with slightly different definition).

Standard error (SE)

1. A statistical measure of an estimate's reliability (that is a measure of sampling error).
2. In NZCASS, the standard error is calculate by the jackknife replicate weights method. See Chapter 12 for further detail.
3. The standard error is used to calculate the margin of error (MoE), the relative standard error (RSE) and confidence intervals.
4. The standard error is also used in tests of statistical significance.

Statistical significance

1. Statistical tests that assess how reliable survey differences are.
2. Because survey estimates are subject to sampling error, differences between years or groups can happen by chance.
3. Statistical significance tests are used to identify whether any differences are 'real' differences between years or groups.
4. Tests at the 95% confidence level are used for the 2014 NZCASS reporting.

Stranger (violent offences by strangers)

1. Violent offences by strangers (sometimes called 'stranger violence') are a subset of 'interpersonal violence' and refer to violence by someone the victim does not know.
2. Within the context of the 2014 NZCASS, 'violent offences by strangers' include a range of offences. See violent interpersonal offences.

Survey (sample survey, victimisation survey)

1. Sample surveys, such as the NZCASS, are a cost-effective alternative to conducting a census.
2. In a survey, a portion of the population is sampled and interviewed to collect the information needed. This contrasts with a census where information is collected from the entire population.
3. Because only a portion of the population is sampled in a survey, published results are considered estimates and there will always be some level of sampling error.

Temporary private dwelling

1. People living in temporary private dwellings are not included in the NZCASS.
2. Temporary private dwellings include caravans, cabins or tents in a motor camp, or boats.

Theft of personal property (general)

1. Property stolen that belongs to the person. This is considered a personal offence, as opposed to burglary, which affects the household.
2. Property stolen when away from home (at the office, place of work, places of entertainment etc).
3. Property stolen in the course of crimes against other people (for example, if property was stolen from a friend's car or mother's house).

Theft from inside home (right to be there)

1. Theft from a building where the person has been invited into that building or has a right to be there.
2. Distinct from 'burglary' as the person has authority to be inside the building.

Theft from motor vehicles

The theft of parts, accessories and other contents from a motor vehicle, such as car stereos, hubcaps and personal items inside the car (like clothing).

Theft from outside the home (over $10)

1. Theft from the yard of a home where the person has been invited into that home or has a right to be there.
2. Theft of household or personal property from the area immediately surrounding the dwelling (but not within an 'enclosed yard').

Theft from person

1. Theft *from a person's body*, such as pickpocketing.
2. Minimal force was used and the offender did not threaten the respondent.

Thefts and damage offences

1. Includes theft from the person, theft of personal property, theft of household property and damage to household or personal property. For the damage offences, offences are only included where the victim did not have contact with the offender or, if the victim was given information on who the offender was, did not know them well.
2. See the NZCASS data items list for more information.

Threats

1. Threats include: verbal or physical threats to kill, injure or assault the respondent and threats to damage personal or household property.
2. Verbal abuse was not counted as a threat.

Threats and damage offences

1. Includes threat of force, threat to damage property, or threat to damage household or personal property. For the damage offences, offences are only included where the victim had contact with the offender or, if the victim was given information on who the offender was, knew them well.
2. Includes interpersonal violence offences, along with sexual offences and physical offences.
3. See the NZCASS data items list for more information.

Urbanisation

1. A classification designed to identify concentrated urban or semi -urban settlements.
2. The NZCASS uses Statistics NZ's standard classification.
3. See the NZCASS data items list for more information.

Usually resident

1. Respondents defined 'usually resident' themselves.
2. Exceptions to self-definition are explained in the Statistics NZ's "Statistical standard for usual residence". See the Statistics New Zealand definition for more information.

Vandalism/damage

1. Wilful damage to personal or household property.
2. Includes (but is not limited to) damage such as arson and graffiti.
3. Excludes incidents that are only a nuisance (like letting down car tyres)

Vehicle

1. Covers all motorised land transport and includes cars, vans, trucks, motorcycles, motor scooters.
2. Can also include vehicles like quad bikes and tractors where they are not out of scope for the NZCASS.

Vehicle offences

1. Includes the four offences related to vehicles: theft of vehicle, damage of vehicle, theft from vehicle, vehicle interference.
2. See the NZCASS data items list for more information.

Vehicle interference

Where someone:

a. tampers with parts of a vehicle

b. tries to get inside a vehicle for an uncertain reason (not a clear attempted theft)

c. successfully gets inside a vehicle but doesn't damage vehicle or intend to do anything else.

Victims

1. Survey respondents who reported experiencing at least one offence against themselves personally or against their household.
2. Victims are counted once for each type of offence experienced, regardless of the number of offences of that type.

Victim forms

1. The part of the survey used to collect detailed information about incidents reported in the survey's screener questions.
2. A respondent can complete a maximum of six victim forms. Respondents could complete three general victim forms at most for incidents reported in the CAPI screener questions and at most one specific victim form for each of the CASI sections.
3. For more information about victim forms see Chapter 3.

Violent interpersonal offences

1. Violent interpersonal offences include: assault offences, kidnapping, robbery, sexual offences, threats and some damage to personal or household property.
2. In the 2014 NZCASS, 'violent interpersonal offences' are discussed as part of analysis and reporting on interpersonal violence. See Interpersonal violence defined for more information.

Weighting/weighted data

1. Weighting is a process of adjusting results from a sample survey to correct for imbalances caused by sampling and survey design.
2. There are three types of weights applied in the NZCASS: household weights, personal weights and incident weights.
3. A weight is allocated to each sample unit, which indicates how many population units are represented by the sample unit.
4. See Chapter 10 for more information.

Acronym List

ABB	approximate Bayesian bootstrap
ANZSOC	Australian and New Zealand Standard Offence Classification
CAPI	computer-assisted personal interviewing
CASI	computer-assisted self-interviewing
CBG	CBG Public Sector Surveying
CI	confidence interval
IQA	independent quality assurance
MoB	month of birth
MoE	margin of error
NAU	Nielsen Area Unit
NIA	National Intelligence Application
NZCASS	the New Zealand Crime and Safety Survey
PAF	Postal Address File
PPS	probability proportional to size
PRINCE2	Projects in a Controlled Environment
PSU	primary sampling unit
RCVS	recorded crime victim statistics
RSE	relative standard error
SC	self-completion
TSS	The Survey System
VF	victim form