Statistics New Zealand
TATAURANGA AOTEAROA

User guide for the 2012 New Zealand General Social Survey confidentialised unit record file

New Zealand Government

Citation
Statistics New Zealand (2013). *User guide for the 2012 New Zealand General Social Survey confidentialised unit record file.* Wellington: Statistics New Zealand

ISBN 978-0-478-40848-5 (online)

Published in September 2013 by
Statistics New Zealand
Tatauranga Aotearoa
Wellington, New Zealand

Contact
Statistics New Zealand Information Centre: info@stats.govt.nz
Phone toll-free 0508 525 525
Phone international +64 4 931 4610
www.stats.govt.nz

Contents

List of tables

1 Introduction

This guide provides information for researchers about confidentialised unit record data from the 2012 New Zealand General Social Survey (NZGSS). Data from this survey is available as a confidentialised unit record file (CURF), which is released with the approval of the Deputy Government Statistician on behalf of the Government Statistician.

Once the application is approved, CURFs can be downloaded via secure Internet access or they can be supplied on a CD-ROM. The guide contains:

- details about what data is included in the 2012 NZGSS CURF
- a summary of the methods we used to confidentialise the data and the differences between the original survey data and the CURF data. This will help you assess the likely effect of the data changes on your analyses
- details on how we calculated replicate weights and how you can use them to calculate standard errors
- details of the contract you need to sign to use the CURF, and which outlines the conditions of release.

About the NZ General Social Survey

The objectives of the New Zealand General Social Survey (NZGSS) are to:

- complement other measures of societal progress by providing information on the well-being of New Zealanders aged 15 years and over across a range of aspects of life
- provide a view of how well-being varies across different groups within the population
- understand the relationships between different aspects of life and to overall well-being.

The survey helps to identify key well-being issues, enables international comparisons to be made, and contributes to better-informed public debate about how we are faring.

The NZGSS 2012 is the third survey in the series. The first NZGSS was carried out in 2008/09 and the information released in October 2009. The second NZGSS was carried out in 2010/11 and the information released in November 2011.

See the 'Definitions' and ‘Data quality’ sections of New Zealand General Social Survey: 2012 (Statistics NZ, 2013), or email info@stats.govt.nz for more technical information about the NZGSS.

About the NZ General Social Survey microdata

CURFs are unit record data that has been modified to protect respondents’ information while maintaining the integrity of the data. The CURF protects respondents’ information against list matching and recognition. By confidentialising microdata before it is released, Statistics NZ helps prevent any intentional or unintentional disclosure of respondent information.

The 2012 NZGSS CURF contains 8,462 individual records. Subject to the limitations of the sample size and the data classifications used, it is possible to manipulate the data, produce tabulations and undertake statistical analyses to individual specifications.

See NZGSS Data Dictionary (Statistics NZ, nd, b) for a list of the data items included in the CURF.

Unit record data from the NZGSS provides valuable information for the research community. If you want data at a finer level of detail for an approved research project, please apply to use more detailed microdata from Statistics NZ's Data Lab. You can use this data on-site at the Data Lab.

Feedback from researchers is always welcome. Please email the CURF Programme at access2microdata@stats.govt.nz with comments and ideas for improving the file.

2 Using the data

The NZGSS CURF contains the following files:

- confidentialised survey data in SAS format
- replicate person weights data in SAS format
- confidentialised survey data in comma separated value (CSV) format
- replicate person weights data in CSV format
- user guide in PDF format
- SAS formats to apply descriptors to the CURF data
- NZGSS questionnaire flowcharts in PDF format
- NZGSS Data Dictionary in Microsoft Excel format.

See the NZGSS Data Dictionary (Statistics NZ, nd, b) for a list of data items included in the NZGSS CURF and the valid ranges/codes for each variable.

Using flowcharts of the questionnaire design

We include flowcharts to show the questions used in the 2012 survey and the questionnaire routing. This will help you determine whether questions are answered by all survey respondents or a subgroup of respondents. By understanding the questionnaire routing, you can choose the appropriate denominator for your analysis.

For example, respondents are asked “In the last 12 months, have you been treated unfairly or had something nasty done to you because of the group you belong to or seem to belong to?” (HUMQ05). Respondents who answer ‘yes’ to this question are asked about the situation they were in when they experienced discrimination (HUMQ07 series) and the reasons for discrimination (HUMS1Q1 series).

If you were analysing the data from these questions, you might calculate the percentage of people who experienced discrimination (HUMQ05 as a percentage of total respondents) then analyse the situations where discrimination was experienced for the subgroup of people who indicated they had experienced discrimination (HUMQ07 as a percentage of people who answered ‘yes’ to HUMQ05).

If you need help understanding the flowcharts, please email info@stats.govt.nz.

Using weights

As the NZGSS is a sample survey, you should apply weights to your analysis. The survey was conducted on a sample of private dwellings in New Zealand. Each record represents a number of people in the population. The number of people represented by a record is indicated by the weight held on the record and is called the final weight (PersonGSSFinalweight). However, it should be noted that as a result of some of the changes made to protect the CURF’s confidentiality, the population estimates generated by using the weights on the CURF may differ slightly from those previously published.

3 Confidentiality methods

To fulfil the obligations of confidentiality under the Statistics Act 1975, the following modifications were made to the NZGSS data:

- collapsing/re-categorisation (global recoding)
- dropping variables
- special uniques analysis (local recoding)
- rounding weights.

Collapsed/recategorised variables (global recoding)

Response options of ‘don't know’ and ‘refused’ are collapsed into the category ‘don't know/refused’.

The NZGSS Data Dictionary shows how response options are collapsed. Other collapsing of categories is shown in appendix 1 of this guide.

Dropping variables

We removed variables that were considered too risky for release, or had a very low response level, from the CURF dataset. This includes variables that:

- input into derived variables (DVs) – if a variable is used in creating a DV then, in general, it is not included in the CURF. For example, in the paid work module, respondents were asked how many hours they worked in their first job, second job (if they had one), third, fourth, and so on. The derived variable ‘WORDV2’ calculates ‘total usual hours worked’ using all this information. The variables with information about hours worked for individual jobs are excluded from the CURF. See the NZGSS Data Dictionary for a comprehensive list of all such occurrences
- were considered a confidentiality risk: ‘regional council’ and ‘household composition’ are excluded for confidentiality reasons. Also, a range of 'human rights' variables (HUMS1Q01_11_R11 to HUMS1Q01OthNR_R25) are excluded because of low frequencies (however, the reason and situation for discrimination are still available in variables that combine both pieces of information). See the NZGSS Data Dictionary for a comprehensive list of all such variables
- contain text responses: certain variables in the NZGSS contain text responses. This is most commonly when the respondent answers ‘other’ to a question. These variables are excluded for confidentiality reasons.

Special unique analysis (local recoding)

Combinations of selected variables were used to determine special unique records. The selected variables used were region, sex, sources of income, ethnic group (single/combination), age, family type, labour force status, social marital status, urban area, and country of birth. Special unique records have either had their value changed to a residual category or had their region changed.

The effect of special unique analysis on the CURF is that 266 cells had their values changed. This equates to 255 records, approximately 3 percent of the sample population.

Rounding

Weights are rounded up or down to integer values.

4 Reconciling data

For some variables, you will not be able to reconcile exactly the data produced from the CURF with already published NZGSS survey results (Statistics NZ, 2013), due to steps we took to preserve confidentiality. The steps were as follows:

- Some variables involved in the special unique analysis process had some responses shifted to residual categories. The variables were: family type, number of children in family, sources of income, urban area, and country of birth. Consequently, estimates produced by the CURF for these variables may have smaller values.
- The special unique analysis applied to the CURF (see chapter 3) perturbed the region of some unit records. See table 1 for examples of differences between published values and corresponding values produced by the CURF.
- Weights are rounded to integer values. The published data in New Zealand General Social Survey: 2012 contains percentages calculated from unrounded weights. Therefore estimates you produce using the CURF may differ marginally from the published data.

Table 1

Example of differences between published and CURF values

Adequacy of money to meet everyday needs being 'enough'

By region

	Adequacy to meet everyday needs being 'enough'		
	Published NZGSS results	CURF	Difference
Region	Percent		Percentage points
Auckland	32.5	30.8	1.7
Wellington	38.0	36.3	1.7
Northland / Bay of Plenty / Gisborne	37.6	35.3	2.3
Rest of North Island	37.8	35.8	2.0
Canterbury	41.4	39.3	2.1
Rest of South Island	38.7	36.6	2.1

5 Sampling error

The 2012 NZGSS used a complex multi-stage sampling design. This means it is not possible for you to use many of the standard methods for calculating standard errors[1] for estimates without having additional information about the sample design.

For you to be able to use these methods, we would need to add primary sampling unit (PSU) and stratum indicator variables to the NZGSS CURF. However, there is an underlying risk with supplying PSU/stratum indicators, because of extra information they might reveal about detailed geography.

As an alternative, we added rounded replicate weights to the NZGSS CURF, to enable you to calculate standard errors.

Model of relative sampling error

We modelled the likely relative sampling errors of estimates (based on 95 percent confidence limits) of different sizes for the 2012 NZGSS (see table 2).

To establish the relative sampling error of a particular number, find the closest figure to the estimate of interest (in the left-hand column) and read off the corresponding relative sampling error (right-hand column). For example, a total estimate of 10,500 people would have relative sampling errors between 32.6 and 43.9 percent, and hence, these estimates should be used with caution.

Table 2

2012 NZGSS sampling error estimates

Estimate	Relative sampling error (Percent)
1,000	118.5
2,000	87.9
3,000	73.8
4,000	65.2
5,000	59.2
10,000	43.9
20,000	32.6
30,000	27.4
40,000	24.2
50,000	22.0
100,000	16.3
200,000	12.1
300,000	10.1
400,000	9.0
500,000	8.1
600,000	7.5
700,000	7.0
800,000	6.6
900,000	6.3
1,000,000	6.0

[1] Sampling error = 1.96 x standard error

What replicate weights are and why we use them

Statistics NZ's household surveys are complex surveys with a multi-stage design. To allow for the effect of this design on the standard errors, Statistics NZ uses a jackknife replication method to produce replicate weights. Replicate weights are a series of variables that contain the information necessary for data users to calculate standard error for statistics such as totals, means, and proportions.

In the NZGSS, we produce the replicate weights using Kott's (2001) delete-a-group jackknife method. PSUs are randomly sorted within each stratum, and then systematically allocated into 100 groups with a constant skip interval. One hundred replicate samples are then formed by deleting one of these 100 groups, and replicate weights are calculated. Each respondent ends up with 100 replicate person weights (PersonGSSFinalWeight_1 to 100).

The fundamental idea behind replicate weights is to use them to calculate statistics in each replicate sample. We estimate the standard error of the full sample by using the variability among the replicate statistics calculated from the replicate samples. You can use the replicate weights for calculating standard errors for many statistics, including totals and proportions. However, for medians or quartiles, a more complex method is required.

Compared with the traditional analytic formulae used to calculate standard errors, the jackknife replication method is much easier to use and more flexible. Separate standard error formulae do not need to be derived for each new complex estimate. Standard errors can be calculated for regression coefficients.

Using the replicate weights we can, for example, also produce quantities required for adjusting chi square tests for the effect of the complex design (see Skinner, Holt, & Smith, 1989, chapter 4).

Final weights and replicate weights in the 2012 NZGSS CURF

There is one final person weight (PersonGSSFinalweight) in the 2012 NZGSS CURF, which is used to calculate estimates. There are 100 replicate person weights (PersonGSSFinalWeight_1 to _100), which are used to calculate the standard error. These are available in a separate file.

See appendix 3 for an example of a simple SAS program you could use to calculate the variance and sampling error for the total of overall life satisfaction by sex, and a logistic regression of overall life satisfaction by age group.

How to use the replicate weights to calculate sampling errors

This is how to use the replicate weights in the NZGSS CURF.

1. Calculate the required estimate using the final weight from the full sample (PersonGSSFinalweight).
2. Calculate the same estimate for each replicate weight (weight1 – weight100).
3. Calculate the variance and sampling error associated with the estimate using the following formula:

The estimate of the variance of $\hat{y}$ is:

$$V(\hat{y}) = \frac{G-1}{G}\sum_{g=1}^{G}\left[\left(\hat{y}_{(g)} - \hat{y}\right)^2\right]$$

and the sampling error at the 95 percent confidence level is:

$$SE(\hat{y}) = \sqrt{V(\hat{y})} \times 1.96$$

where:

G = the total number of replicates (in this case 100)

g = the replicate

$\hat{y}$ = a population estimate from the wgt0 (full sample weight)

$\hat{y}_{(g)}$ = a replicate population estimate (from wgt1 ==> wgt100)

When to use the replicate weighting approach

You can use the replicate weighting approach to calculate the standard error of most statistics such as means, percentages, ratios, correlations, derived statistics, and regression coefficients.

Software packages for computing replicate standard errors and variances

WesVar (Westat 2007) is a common complex-survey analysis software package that computes sampling errors and variance estimates using replicate weights (version 4 will also calculate sampling errors for percentiles). See reference to Westat, (2007) for documentation, including PDF.

Other statistical analysis software packages include SUDAAN and VPLX.

6 Conditions of release

According to the Statistics Act 1975, the Government Statistician may release the 2012 NZGSS CURF to approved users under section 37(4):

> All statistical information published is to be arranged in such a manner as to prevent any particulars published from being identifiable by any person as particulars relating to any particular person or undertaking.

Researchers from the public and private sectors within New Zealand and Australia may apply for a CURF, provided the file is used for statistical purposes (see Confidentialised unit record files, Statistics NZ nd, a).

The security and ethical terms and conditions to be followed are clearly outlined in the documents issued to CURF applicants: CURF Licence Agreement between Statistics NZ and the institution, CURF Cost and Cancellation Policies, and this guide.

This guide is the researcher's primary information source for the security and ethical obligations of participating in the CURF programme. However, do not overlook the importance of the other two documents.

Here is an overview of the terms and conditions, as stated in the undertaking for the researcher. The researcher states that (s)he will:

- use the CURF only for the statistical purposes specified in schedule A of the signed research undertaking
- not attempt to identify particular persons or organisations
- not disclose, either directly or indirectly, any individual's information on the CURF to any other person, organisation, or participating researchers of this undertaking
- not attempt to match the information, with or without using identifiers, with any other unit record level data source
- not attempt to access the information after the term of research has completed or the CURF authorisation expires and the researcher(s) have chosen not to renew
- store the CURF in a locked cabinet and in a password secure location (computer). This CURF is to be accessed only by the researcher(s) who have signed this undertaking for this specific project
- properly destroy all identifiable printouts as confidential waste when no longer required
- keep all readable printout of unit records on the organisation’s premises and in a locked cabinet when not in use
- allow Statistics NZ to inspect the security system established by the researcher(s) if they ask
- use replicate weights only to calculate the estimate of standard error
- summarise research and CURF usage of the past year if study extends beyond one year and every year after
- return the CURF when the research is completed or when the end date stated in the CURF Licence Agreement between Statistics NZ and the institution has passed.

Researchers must first read this guide, then select a responsible officer to oversee the project, and finally, complete the application to access microdata form for CURFs on the Statistics NZ website. Once a responsible officer is identified, this individual is accountable for reading this guide so they understand their role.

There are three CURF application documents that require either the researcher's or the responsible officer's signatures. Once signed, these documents are a binding agreement between the researcher, the responsible officer, and Statistics NZ. They agree to comply with the terms and conditions of the CURF programme.

7 References

Kott, P (2001). The delete-a-group jackknife. *Journal of Official Statistics, 17*(4), 521–526.

Skinner, CJ, Holt, D, & Smith, TMF (1989). *Analysis of complex surveys.* Indianapolis: Wiley & Sons Inc.

Smith, H. (2001). Investigation of the delete-a-group jackknife variance estimator for the HLFS. Wellington: Statistics New Zealand. Available from www.stats.govt.nz.

Statistics New Zealand (nd, a). Confidentialised unit record files. Available from www.stats.govt.nz.

Statistics New Zealand (nd, b). New Zealand General Social Survey Data Dictionary. Available from www.stats.govt.nz/nzgss.

Statistics New Zealand (2009). New Zealand General Social Survey: 2008.

Available from www.stats.govt.nz/nzgss.

Statistics New Zealand (2011). New Zealand General Social Survey: 2010.

Available from www.stats.govt.nz/nzgss.

Statistics New Zealand (2013). New Zealand General Social Survey: 2012.

Available from www.stats.govt.nz/nzgss.

Westat (2007). WesVar® 4.3 User's guide. Rockville, MD: Westat. Available from www.westat.com.

Appendix 1 – List of variables

The following table shows how the New Zealand General Social Survey variables are modified (collapsed) in the 2012 NZGSS CURF. The table also shows changes in collapsing between the 2012 and 2010 CURFs.

A complete list of all the variables in the 2012 NZGSS CURF is in the Excel file 'NZGSS Data Dictionary.xls', which is available on the Statistics NZ website at New Zealand General Social Survey Data Dictionary.

Note this table does not include changes to the actual NZGSS questions. See the data dictionary for more details of variables in the 2012 dataset and CURF.

SAS name	Description	Range	2012 CURF range	2010 CURF range
CORPQ13	What year did you first arrive in New Zealand to live?	00 – text response (individual years)	Pre-1960 1961 1962 … 2009 2010 2011 2012–13	Pre-1945 then individual years
CORPQ17	In what year did you complete this qualification?	00 – text response (individual years)	01. 2005 to 2013 02. 2000 to 2004 03. 1995 to 1999 04. 1990 to 1994 05. 1985 to 1989 06. 1980 to 1984 07. 1975 to 1979 08. 1970 to 1974 09. 1965 to 1969 10. 1960 to 1964 11. 1955 to 1959 12. 1950 to 1954 13. 1945 to 1949 14. 1940 to 1944 15. before 1940	01. 2005 to 2011 02. 2000 to 2004 03. 1995 to 1999 04. 1990 to 1994 05. 1985 to 1989 06. 1980 to 1984 07. 1975 to 1979 08. 1970 to 1974 09. 1965 to 1969 10. 1960 to 1964 11. 1955 to 1959 12. 1950 to 1954 13. 1945 to 1949 14. 1940 to 1944 15. before 1940

Table continued next page

Table continued

SAS name	Description	Range	2012 CURF range	2010 CURF range
CORPQ21	In what year did you complete this qualification?	00 – text response (individual years)	01. 2005 to 2013 02. 2000 to 2004 03. 1995 to 1999 04. 1990 to 1994 05. 1985 to 1989 06. 1980 to 1984 07. 1975 to 1979 08. 1970 to 1974 09. 1965 to 1969 10. 1960 to 1964 11. 1955 to 1959 12. 1950 to 1954 13. before 1950	01 – 2005 to 2011 02 – 2000 to 2004 03 – 1995 to 1999 04 – 1990 to 1994 05 – 1985 to 1989 06 – 1980 to 1984 07 – 1975 to 1979 08 – 1970 to 1974 09 – 1965 to 1969 10 – 1960 to 1964 11 – 1955 to 1959 12 – 1950 to 1954 13 – Before 1950
CORDV5	Number of dependent children in family	01 – one dependent child 02 – two dependent children 03 – three dependent children 04 – four or more dependent children 05 – no dependent children	01 – one dependent child 02 – two dependent children 03 – three or more dependent children 05 – no dependent children 77. number of children not identifiable	
CORDV8_31	Sources of personal income: wages, salary, commissions, bonuses etc. paid by an employer	0. no 1. yes	0. no 1. yes	(NAMED CORDV8_11 in GSS2 CURF and in original dataset)
CORDV8_32	Sources of personal income: self-employment or business	0. no 1. yes	0. no 1. yes	(NAMED CORDV8_12 in GSS2 CURF and in original dataset)
CORDV8_33	Sources of personal income: interest, dividends, rent, other investments	0. no 1. yes	0. no 1. yes	(NAMED CORDV8_13 in GSS2 CURF and in original dataset)
CORDV8_34	Sources of personal income: NZ Superannuation or veteran's pension	0. no 1. yes	0. no 1. yes	(NAMED CORDV8_14 in GSS2 CURF; CORDV8_15 in original dataset)

Table continued next page

Table continued

SAS name	Description	Range	2012 CURF range	2010 CURF range
CORDV8_35	Sources of personal income: other superannuation pensions, annuities (other than NZ Superannuation, veteran's pension or war pension)	0. no 1. yes	0. no 1. yes	NAMED CORDV8_15 in GSS2 CURF; CORDV8_16 in original dataset
CORDV8_36	Sources of personal income: government benefits	CORDV8_17 CORDV8_18 CORDV8_19 CORDV8_20.	0. no 1. yes	NAMED CORDV8_16 in GSS2 CURF
CORDV8_37	Sources of personal income: other government transfers	CORDV8_14 CORDV8_21 CORDV8_22	0. no 1. yes	NAMED CORDV8_17 in GSS2 CURF
CORDV8_38	Sources of personal income: other sources of income	0. no 1. yes	0. no 1. yes	NAMED CORDV8_18 in GSS2 CURF and in original dataset
CORDV8_39	Sources of personal income: no source of income during that time	0. no 1. yes	0. no 1. yes	NAMED CORDV8_19 in GSS2 CURF and in original dataset
UA	Urban area	Currently excluded	01. main urban 02. secondary urban 03. minor urban 04. rural 77. urban area not identifiable	01. main urban 02. secondary urban 03. minor urban 04. rural
KASQ02OthNR	Non-response to other reason for feeling dissatisfied with knowledge, skills, and abilities	88. don't know 99. refused	excluded	89. DK/RF
KASQ05_88	Person doesn't know what prevented them from doing the study or training they would like to do	0. no 1. yes	excluded	KASQ05_88 & KASQ05_99 are combined into KASQ05_89
KASQ05_99	Person refuses to answer what prevented them from doing the study or training they would like to do	0. no 1. yes	excluded	KASQ05_88 & KASQ05_99 are combined into KASQ05_89
WORQ06NR	Non-response to tasks and duties in their main job	88. don't know 99. refused	excluded	89. DK/RF

Table continued next page

Table continued

SAS name	Description	Range	2012 CURF range	2010 CURF range
WORDV2	Total usual hours worked derived variable	000. (0) total usual hours worked 001. (1) total usual hours worked … xxx. (xx) total usual hours worked 777. not stated	01. 1 – 5 hours 02. 6 – 10 hours 03. 11 – 15 hours 04. 16 – 20 hours 05. 21 – 25 hours 06. 26 – 30 hours 07. 31 – 35 hours 08. 36 – 40 hours 09. 41 – 45 hours 10. 46 – 50 hours 11. 51 – 55 hours 12. 56 – 60 hours 13. 61 – 70 hours 14. 71 – 80 hours 15. 81 hours and over 99. Not stated	000 hours 001 hours 002 hours … 059 hours 060 hours 61–70 hours 71–80 hours 81 hours and over
HOUQ03_19	Person doesn't have any of the major problems (listed on showcard) with their house/flat	0. no 1. yes	0. no 1. yes	
HOUQ03_89	Person doesn't know OR refuses to answer what the major problems with their house/flat are	HOUQ03_88 HOUQ03_99	HOUQ03_88 HOUQ03_99	
HOUQ04_17	Barking dogs is a major problem with the person's street/neighbourhood	0. no 1. yes	0. no 1. yes	
HOUQ04_18	Person has other problems with their street/neighbourhood which are not listed on the showcard	0. no 1. yes	0. no 1. yes	
HOUQ04_19	Person doesn't have any of the major problems (listed on showcard) with their street/neighbourhood	0. no 1. yes	0. no 1. yes	
SAFQ05	How many of those crimes involved violence?	11. none 12. one 13. more than one 88. don't know 99. refused	11. none 12. one 13. more than one	01. yes 02. no 89. DK/RF

Table continued next page

Table continued

SAS name	Description	Range	2012 CURF range	2010 CURF range
SAFQ10_89	Person doesn't know OR refuses to answer what happened to them as a result of non-violent crime(s) in the last 12 months	0. no 1. yes	excluded	SAFQ10_88 & SAFQ10_99 are combined into SAFQ10_89
SAFQ19_89	Person doesn't know OR refuses to answer what happened to them as a result of traffic accident(s) in the last 12 months	0. no 1. yes	excluded	SAFQ19_88 & SAFQ19_99 are combined into SAFQ19_89
SUPQ09_89	Person doesn't know OR refuses to answer if they (or their partner) give support to family or relatives aged under 18 who don't live with them	0. no 1. yes	excluded	SUPQ09_88 & SUPQ09_99 are combined into SUPQ09_89
SUPQ11_89	Person doesn't know OR refuses to answer if they (or their partner) give support to family or relatives aged 18 to 24 who don't live with them	0. no 1. yes	excluded	SUPQ11_88 & SUPQ11_99 are combined into SUPQ11_89
SUPQ13_89	Person doesn't know OR refuses to answer if they (or their partner) give support to family or relatives aged 25 to 64 who don't live with them	0. no 1. yes	excluded	SUPQ13_88 & SUPQ13_99 are combined into SUPQ13_89
SUPQ15_89	Person doesn't know OR refuses to answer if they (or their partner) give support to family or relatives aged 65 or over who don't live with them	0. no 1. yes	excluded	SUPQ15_88 & SUPQ15_99 are combined into SUPQ15_89
SOCQ12_89	Person doesn't know OR refuses to answer what makes it hard to have contact with family or friends	0. no 1. yes	excluded	SOCQ12_88 & SOCQ12_99 are combined into SOCQ12_89
HUMQ07_89	Person doesn't know OR refuses to answer what situation they were in when they experienced discrimination	0. no 1. yes	excluded	HUMQ07_88 & HUMQ07_99 are combined into HUMQ07_89

Appendix 2 – Technical notes

These technical notes provide information about the 2012 NZGSS and survey population.

See New Zealand General Social Survey: 2012 data quality section for more information.

Target population

The target population for the NZGSS is the usually resident population aged 15 years and over in private dwellings in the North Island, South Island, or Waiheke Island of New Zealand.

The target population includes:

- New Zealand usual residents temporarily overseas
- New Zealand usual residents temporarily staying elsewhere in New Zealand (including other permanent and temporary private dwellings, institutions, and non-private dwellings; and people who have no fixed abode, but are found in private dwellings on the household enumeration date)
- people in the New Zealand armed forces if they reside in a private dwelling
- young adults at boarding schools (young adults who fall into this category are not surveyed in the personal questionnaire, but are included as members of the household in the household questionnaire).

The target population excludes:

- overseas visitors and international students who expect to be resident in New Zealand for less than 12 months
- people living in non-private dwellings such as hotels, motels, boarding houses, hostels, and homes for the elderly
- patients in hospitals, or residents of psychiatric and penal institutions
- people living on offshore islands (excluding Waiheke Island)
- members of the non-New Zealand armed forces and their dependants
- non-New Zealand diplomats or diplomatic staff members and their dependants
- New Zealand usual residents temporarily overseas who do not return within the survey period
- New Zealand usual residents temporarily staying elsewhere in New Zealand (including other permanent and temporary private dwellings, institutions, and non-private dwellings; and people who have no fixed abode, but stay at private dwellings) who don't return within the survey period
- New Zealand usual residents who live in remote areas that are costly or difficult to access.

Response rate

The target response rate for the survey is 80 percent. The achieved response rate for 2012 was 78 percent. The response rate was calculated by dividing the weighted percentage of eligible individuals who responded by the estimated number of eligible individuals.

Usage and limitations of the data

The variables NZDep (NZDep2006) and urban area (UA) should be used with caution due to the age of the classifications.

The NZDep variable in the 2012 NZGSS CURF was created from 2006 Census data and it represents area-based deprivation at the time of the 2006 Census. As the index is more than five years old, it may not reflect the current level of deprivation for the area.

Urban areas are currently defined on the basis of the 1996 Census usually resident population count. The classification is becoming less relevant over time.

Sampling error method

Replicate weights are used for the sampling error estimations. The NZGSS uses the extended delete-a-group (extended DAG) jackknife method for estimating sampling error. This method is commonly used in Statistics NZ and does not require an analytic equation to estimate the sampling errors. It can be used for all estimates except quantiles (ie the totals or proportions that might be the most likely estimates from NZGSS). This method overcomes the upward bias in the variance estimation that may occur in the standard DAG when there are fewer sampled PSUs in a stratum than there are replicate groups being formed. Quantile sampling errors, if required, can be calculated using the Woodruff Method.

Based on an investigation conducted for the Household Labour Force Survey (Smith, 2001), and standard practice among several other household surveys using the extended DAG jackknife method, it was decided to use 100 replicate groups.

The extended DAG jackknife drops a group of PSUs at a time. The PSUs are randomly sorted within a stratum and are then allocated in sequence to one of the G (ie 100) replicate groups. Those PSUs selected in a particular replicate group are then assigned a small weight for that replicate sample. All remaining units get reweighted to compensate for this. The purpose of this is to create 100 different survey estimates based on dropping one percent of the sample each time, and the variance in these estimates is used to calculate the sampling error.

Revisions to sample

To improve the accuracy of the NZGSS we made some changes to the NZGSS 2012 sample, and reissued weights for NZGSS 2010 and NZGSS 2008.

We found the NZGSS 2008 and NZGSS 2010 samples did not have the expected socio-economic dimensions. This was discovered through comparison with the NZ Deprivation Index and is a consequence of non-random sampling.

The effect of this on most published estimates from the NZGSS 2010 and NZGSS 2008 is not statistically significant and it is unlikely to affect aggregate released data. However, point estimates were consistently away from the expected results for lower socio-economic groups, including the Pacific population.

We re-selected a sample for the NZGSS 2012 to ensure the socio-economic bias is corrected. This means the sample has been selected from different primary sampling units than for NZGSS 2010 and NZGSS 2008.

We also reissued the weights for NZGSS 2010 and NZGSS 2008 to improve small group analysis, which we see as an important aspect of the survey.

Appendix 3 – SAS example program

Here is a simple SAS example program that calculates the variance and sampling error for the total of overall life satisfaction by sex, and a logistic regression of overall life satisfaction by age group. You can modify this programme to calculate variance and sampling error for your own analyses.

```
****************************

****************************

** Example SAS program demonstrating how to calculate sampling errors using the
jacknife method

** Two types of estimates are demonstrated - (a) totals, (b) parameters from logistic

** regression

**

** Input: GSS2012curf - a unit record dataset that includes the final person weight
(PersonGSSFinalWeight) and 100 replicate weights

*********************************************

*********************************************;

** Estimate the sampling error for totals using the jacknife method

**********************;

* Specify groups to form the totals by;

%let bygroups = CORDV10 OLSQ01; *e.g. sex by Overall Life Satisfaction;

/*get the weighted total for each group */

proc summary data=GSS2012curf nway;

  var PersonGSSFinalWeight PersonGSSFinalWeight_1 - PersonGSSFinalWeight_100;

  class &bygroups.;

  output out=estimates(drop=_type_ _freq_) sum=sum0 - sum100;

run;

/* calculate the variance and standard error for the total */;

data output (keep=&bygroups. var se rse sum0);

  set estimates;

  format rse percent9.1;

  * Note the value at [0] is the actual (non-replicate) estimate ;

  array sums[0:100] sum0 - sum100;
```

Continued next page

```
  *reset the squared sum of the differences for each new group;
  ss=0;
 * calculate variance and standard error of the estimate using jackknife method;
  do i = 1 to 100;
   ss=ss+(sums[i]-sums[0])**2;
  end;
  var=ss*(100-1)/100;
  se=1.96*var**0.5;
  rse = se / sum0;
run;
*************
** Estimate the sampling error for parameters in logistic regression using the
jacknife method
**************;
%macro regression;
   * Put the regression parameters into a dataset;
   ods output ParameterEstimates=Parameters(keep=estimate variable classval0
      rename = estimate = estimate0);
   ods listing close;
      * Run a regression using the finalweight.
      The standard errors from the regression output should not be used.
      We are running the regression to get the regression parameters and will do
      this for each replicate weight to calculate the sampling errors for each
      parameter;
   PROC LOGISTIC DATA= GSS2012curf ;
      CLASS CORDV9 (PARAM=EFFECT);
      weight PersonGSSFinalweight;
      MODEL OLSQ01=CORDV9     /
         SELECTION=NONE
         LINK=LOGIT;
   RUN;
   QUIT;
```

Continued next page

```
  *Now do the regression for each of the replicate weights, attaching each version
   of the parameters as a new column in the dataset Parameters;
  %do i = 1 %to 100;
      ods output ParameterEstimates=Parms&i.(keep=estimate variable classval0
         rename=estimate=estimate&i.);
      PROC LOGISTIC DATA= GSS2012curf ;
         CLASS CORDV9 (PARAM=EFFECT);
         weight PersonGSSFinalweight_&i.;
         MODEL OLSQ01=CORDV9     /
            SELECTION=NONE
            LINK=LOGIT;
      RUN;
      QUIT;
      proc sort data=parameters; by variable classval0;run;
      proc sort data=parms&i.; by variable classval0;run;
      data parameters;
        merge parameters parms&i.;
          by variable classval0;
      run;
  %end;
%mend;
%regression;
/* Calculate the jackknife variance and standard error for each parameter in a similar
way to the method used for the total */;
data output;*(keep=var se) ;
   set parameters end=eof;
   array par[0:100] estimate0 - estimate100;
   ss=0;
   Do i =1 to 100 ;
      ss=ss+(par[i]-par[0])**2;
   END;
   var=ss*(100-1)/100;
   se=var**0.5*1.96;
run;
```