Scottish Surveys Core Questions 2016

Technical Notes

Office of the Chief Statistician and Strategic Analysis

The Scottish Government

Scottish Surveys Core Questions 2016

The Scottish Surveys Core Questions (SSCQ) 2016 is an annual Official Statistics publication for Scotland. SSCQ provides reliable and detailed information on the composition, characteristics and attitudes of Scottish households and adults across a number of topic areas including equality characteristics, housing, employment and perceptions of health and crime.

The SSCQ gathers survey responses from identical questions in the Scottish Crime and Justice Survey, the Scottish Health Survey and the Scottish Household Survey into one output. The pooling of Core Questions results in an annual sample of around 20,000 respondents, providing unprecedented precision of estimates at national level. This sample size enables the detailed and reliable analysis of key national estimates by country of birth, ethnicity, sexual orientation, religion, age and sex, marital status, education level and economic activity, as well as tenure, car access and household type. SSCQ also enables a detailed sub-national analysis by Local Authority, urban-rural classification and Scottish Index of Multiple Deprivation.

The Scottish Surveys Core Questions in 2016 covered:

- self-assessed general health
- disability and long-term conditions
- smoking
- mental well-being
- provision of unpaid care
- perception of local crime rate
- perceptions of police performance
- highest qualification held
- economic activity
- household type

- housing tenure
- car access
- ethnicity
- religion
- marital status
- sexual orientation
- gender
- age

Jamie Robertson, Ben Cook
- together with numerous colleagues past and present who, along with our contractors, have helped deliver each of the surveys as well as this methodological transformation

Office of the Chief Statistician, The Scottish Government

For more information about SSCQ and the latest publications and outputs, go to www.gov/scot/SSCQ

Contents

Scottish Surveys Core Questions 2016 2
Contents 3
1 Source surveys and core questions 4
1.1 Scottish Crime and Justice Survey (SCJS) technical notes 4
1.2 Scottish Health Survey (SHeS) technical notes 5
1.3 Scottish Household Survey (SHS) technical notes 5
2 Weighting 5
3 Confidence Interval Calculations 6
4 Variable list 8
5 Statistical Disclosure Control 9
5.1 Recoding of Geographical Identifiers 9
5.1.1 HBA – Health Board Areas 9
5.1.2 SIMD5 – Scottish Index of Multiple Deprivation Quintiles 10
5.1.3 UR – Urban Rural 2 fold scale 10
5.2 Recoding of Individual Characteristics 10
5.2.1 Ethnic Group 10
5.2.2 Religion 10
5.3 Survey weighting variables 11
5.4 Cluster variable 11

1 Source surveys and core questions

Results from the three large-scale Scottish Government population surveys are published separately as National Statistics:

- Scottish Crime and Justice Survey (SCJS)
  www.gov.scot/Topics/Statistics/Browse/Crime-Justice/crime-and-justice-survey
- Scottish Health Survey (SHeS)
  www.gov.scot/Topics/Statistics/Browse/Health/scottish-health-survey
- Scottish Household Survey (SHS)
  www.gov.scot/shs

Further information on Population Surveys in Scotland can be found here:
www.gov.scot/Topics/Statistics/About/Surveys

Since the beginning of 2012 each of the surveys has included a set of 20 core questions that provide information on the composition, characteristics and attitudes of Scottish households and adults across a number of topic areas including equality characteristics, housing, employment and perceptions of health and crime. Responses on these questions from all three surveys have been pooled to provide the Scottish Surveys Core Questions (SSCQ) dataset with a sample size of around 20,000 responses.

Full details of the harmonised questions are available on the Scottish Government website[1] and questionnaires are provided on the websites of each of the individual surveys.

The first set of pooled response tables for the year 2012 were published as data under development here: www.gov.scot/Topics/Statistics/About/Surveys/PooledSample2012

Following further consultation and methodological development, the 2013 dataset was published as Official Statistics in December 2015. The website contains further information and supplementary tables to this main report.

www.gov.scot/Topics/Statistics/About/Surveys/SSCQ

The SSCQ 2016 dataset was pooled from the Scottish Crime and Justice Survey 2016/17, Scottish Health Survey 2016 and the Scottish Household Survey 2016. Responses from adults aged 16 and over were included.

Due to the different sampling nature of each survey, which is necessary to meet their primary aims, the number of respondents varies between different SSCQ questions. The questions were hence batched into three groups: household questions, individual questions and crime questions, and three different sets of weights calculated to ensure representative results. Sampling, weighting and pooled sample numbers are described separetely for each survey below.

1.1 Scottish Crime and Justice Survey (SCJS) technical notes

Sampling, survey response and weighting are described in full in the forthcoming SCJS 2016/17 technical report:

[1] http://www.gov.scot/Topics/Statistics/About/SurveyHarm

Briefly, the survey consists of a simple random sample, designed to achieve a minimum effective sample size of 315 interviews in the 14 Police Divisions in Scotland (PDs), enabling robust analysis at this level. One random adult per household was interviewed and asked all SSCQ and SCJS questions.

From 2016/17 onwards, the SCJS is run continuously annually, which means that annual fieldwork periods can be drawn into the pool on a calendar year basis.

Due to errors in SCJS fieldwork in the 2016/17 and 2017/18 sweeps, country of birth was not collecgted in full and has been removed from the core for this period. Remedial work may result in a smaller country of birth sample and an additional survey weight being added in future iterations of this data set.

1.2 Scottish Health Survey (SHeS) technical notes

Sampling, survey response and weighting are described in full in the SHeS 2016 technical report: http://www.gov.scot/Publications/2017/10/4796

The SHeS sample is clustered in each calendar year and unclustered over four years. All adults and up to two children in each household are eligible for interview. Only one adult in each household was asked the crime and household questions, to remain in line with the SCJS sampling procedure. The SHeS sample is boosted by participating health boards. It is further boosted to interview children in further households. These households were excluded from the SSCQ dataset as equality questions were not asked.

The response rate was 58%, and 3,334 households were interviewed in the main and health board boost samples. 4,323 resulting adult interviews were pooled into the SSCQ 2016 dataset. Of these, 2,801 were asked the crime questions. The subset of households (excluding the child boost), and adult respondents were re-weighted to be representative of the Scottish private household and population distribution, as described for the SHeS publication.

1.3 Scottish Household Survey (SHS) technical notes

Sampling, survey response and weighting are described in full in the SHS 2016 technical report.
http://www.gov.scot/Topics/Statistics/16002/PublicationMethodology/Methodology16

The SHS consists of a simple random sample with a target minimum effective sample size of 250 per local authority. The SSCQ household questions are asked of the highest income householder or their spouse/partner, and one adult is randomly selected to answer the individual and crime questions, in line with the other two surveys.

64% of eligible households responded, leading to 10,470 household interviews. The response for the random adult interview was 59%, yielding 9,642 interviews. Weighting is fully described in the SHS technical report.

2 Weighting

Datasets from the three source surveys were combined into three new SSCQ datasets: SSCQ household variables (19,371 responses), SSCQ individual variables (19,532 responses) and SSCQ crime variables (18,010 responses), see Table 1.

Each variable response category in each of the surveys carries a different design effect. If we were solely seeking the most efficient estimate for each variable separately then separate scale factors could be derived for each one. However, this would restrict the use and understanding of the dataset. Rather, for each constituent survey dataset the design effects were estimated for each response category and then the median design effect over all response categories for all variables was used as the representative design effect of that survey. These design effects were then used along with the sample sizes to calculate the effective sample sizes (neff) and scaling factors for combining the three datasets.

Table 1: Numbers of sample and effective sample pooled from the source surveys

	SCJS		SHeS		SHS		SSCQ	
	sample	neff	sample	neff	sample	neff	sample	neff
Household responses[2]	5,567	5,044	3,334	2,116	10,470	8,770	19,371	15,930
Individual responses[3]	5,567	4,204	4,323	2,102	9,642	6,625	19,532	12,931
Crime responses[4]	5,567	4,107	2,801	1,301	9,642	6,474	18,010	11,882

To combine the data the scale factors were applied to the grossing weights for the individual surveys (described in section 1). The neff of each survey contribution formed the basis for the scaling factors:

survey A weight scaling factor = neff (surveyA) / (sum of three survey neffs).

The weights were then re-scaled to be proportionate to effective sample size contribution of each survey and used as pre-weights.

The three pooled SSCQ datasets were then weighted again to be representative of National Records of Scotland population estimates[5].

3 Confidence Interval Calculations

All three of the source surveys are stratified to ensure sufficient sample sizes in the smaller local authority areas. In addition, SHeS is clustered in each annual fieldwork period and, while this effect cancels out over each four-year period, it must be accounted for in producing annual results.

[2] SSCQ household variables are household type, tenure and car access

[3] SSCQ individual respondent variables are self-assessed general health, limiting long-term health conditions, smoking, unpaid care provision, mental wellbeing, highest achieved qualification, economic activity, country of birth, ethnic group, religion, marital status, sexual orientation, sex and age

[4] SSCQ crime variables are perception of local crime rate and six questions on perceptions of police performance

[5] See SSCQ Weighting tables spreadsheet at http://www.gov.scot/Topics/Statistics/About/Surveys/SSCQ/SSCQ2016/WgtBase

Confidence intervals have therefore been calculated using a method to account for stratification and clustering and the resulting design effects (surveyfreq in SAS). This method is used to compare estimates of all quantities provided by SSCQ. Confidence intervals across all subgroup estimates are provided in the accompanying supplementary tables.[6]

Confidence intervals are plotted on all charts and figures in this report. If the intervals do not overlap then there is a significant difference between two points, but if they do overlap it does not necessarily mean there is no significant difference.[7] In the report text the term "significant" refers to "statistically significant" differences.

A comparison of estimates of key variables across the three constituent surveys and the SSCQ are provided in Annex A of the main report[8].

[6] SSCQ Supplementary Tables available at http://www.gov.scot/Topics/Statistics/About/Surveys/SSCQ/SSCQ2016/SuppTables

[7] see guidance at http://www.gov.scot/Topics/Statistics/About/Methodology/confinv

[8] SSCQ Equality Report 2016, http://www.gov.scot/Topics/Statistics/About/Surveys/SSCQ/SSCQ2016/Report

4 Variable list

Group	Variable name	Label
Individual	ageband	Respondent age group
	Sex	Respondent sex
	EthnicGroupBand	Respondent ethnic group
	ReligionBand	Respondent religious group
	MaritalStatusBand	Respondent marital status
	TopQualBand	Respondent highest qualification
	ILOempBand	Respondent employment status
	GenHealthBand	Respondent self-assessed general health
	LTConditionBand	Respondent long-term limiting health condition
	smokingBand	Respondent smokes
	IndCareBand	Respondent provides unpaid care
	swemwbs	Shortened Transformed Warwick-Edinburgh Mental Wellbeing Score
	pooled_ind_wt	Adult grossing weight
	pooled_ind_wt_sc	Adult scaled weight
Household	HType	Household type
	htype2a	Household type
	tenureBand	Household tenure
	outtenBand	Household tenure
	CarAccessBand	Household access to car
	pooled_hh_wt	Household grossing weight
	pooled_hh_wt_sc	Household scaled weight
Crime	CrimeAreaBand	Perception of change in local crime rate
	POLCONA	Conf in police to - prevent crime
	POLCONB	Conf in police to - respond quickly to app calls and inf from public
	POLCONC	Conf in police to - deal with incidents as they occur
	POLCOND	Conf in police to - investigate incidents after they occur
	POLCONE	Conf in police to - solve crimes
	POLCONF	Conf in police to - catch criminals
	pooled_crim_wt	Adult grossing weight
	pooled_crim_wt_sc	Scaled adult weight
Geography	LA	Local authority
	HBA	Community Health Board
	PSD	Police Scotland Division
	UR	Urban/rural indicator
	SIMD5	Scottish Index of Multiple Deprivation: quintile groups
	CLUSTER	Cluster variable

5 Statistical Disclosure Control

GSS Disclosure Control Guidance for Microdata Produced from Social Surveys (September 2014) cites principle 5 of the Code of Practice for Official Statistics for guidance on the publication of microdata under EUL, requiring that: *"official statistics do not reveal the identity of an individual or organisation, or any **private information** relating to them, taking into account other relevant sources of information." "Information relating to an individual should be considered by a producer of statistics to be 'private' if it was provided with the expectation that the information would be kept out of the public domain. Survey pledges provide respondents with assurances that the information they provide will remain confidential."*

Furthermore, the guidance states that "arrangements for confidentiality protection are sufficient to protect the privacy of individual information, but not so restrictive as to limit unduly the practical utility of official statistics".

Private information is defined in the National Statistician's Guidance; Confidentiality of Official Statistics as being information that:

- relates to an identifiable legal[9] or natural person, and
- is not in the public domain or common knowledge, and
- if disclosed would cause them damage, harm or distress

Three methods were used to disclosure control the main data set: re-coding, banding and variable removal.

- Re-coding is used on categorical variables to collapse responses categories together into larger groups to hide small number of respondents in some of the smaller groups.
- Banding is used on numerical variables such as age, where discrete values are grouped sequentially, e.g. where a respondent was aged 80 or over their age was re-coded as '80+'
- Variable removal was used to remove:
    - sensitive variables;
    - variables used to calculated summary variables which can be disclosive if not recoded, or used to help identify respondents in combination with other variables.

5.1 Recoding of Geographical Identifiers

5.1.1 HBA – Health Board Areas

HBA follows the health board area boundaries pre 1st of April 2014 and when combined with LA create small pockets of respondents in very specific geographies. HBA has been recoded to align with LA as per the change to health board boundaries implemented on the 1st of April 2014.

[9] A 'legal person' is a company, enterprise, or other organisation that has a legal identity. A 'natural person' is a member of the public. Where the term 'individual' is used in the Code it means both legal and natural persons, both living and dead.

5.1.2 SIMD5 – Scottish Index of Multiple Deprivation Quintiles

Analysing SIMD quintile data zones, there 7 unique data zones across Scotland and if a respondent in the survey sample is from these data zones it could lead to their identification. As a result SIMD5 has been recoded to adsorb these 7 unique data zones into a neighbouring quintiles within LA.

5.1.3 UR – Urban Rural 2 fold scale

Analysing urban rural two fold data zones in combination with SIMD quintile data zones and LA data zones reveals 36 rare Scottish data zones (< three data zones per variable combination), and if a respondent in the survey sample is from these data zones it could lead to their identification. As a result, UR has been recoded to adsorb these data zones into the rural category for urban responses and vice versa for a rural responses for the 36 unique data zones.

5.2 Recoding of Individual Characteristics

5.2.1 Ethnic Group

Ethnicity of respondents is collected in detail by the contributing surveys to the SSCQ. Ethnic groups have been collapsed into the following categories:

- White - Scottish
- White - British
- White - Other
- Minority Ethnic
- DK / Refused

5.2.2 Religion

Religious groups have been grouped into the following categories:

- None
- Church of Scotland
- Roman Catholic
- Other Christian
- Another Religion / DK

5.3 Survey weighting variables

The following weighting variables were rounded to two decimal places to increase uncertainty in the estimates calculated by users:

- pooled_ind_wt
- pooled_ind_wt_sc
- pooled_hh_wt
- pooled_hh_wt_sc
- pooled_crim_wt
- pooled_crim_wt_sc

5.4 Cluster variable

The variable CLUSTER has been anonymised using a random interger to prevent disclosure. The number is therefore meaningless other than as a grouping variable for clustered sample units.