Scottish Surveys Core Questions 2012

Technical Notes

Office of the
Chief Statistician and Strategic Analysis

The Scottish Government

Scottish Surveys Core Questions 2012

Following the publication of the Scottish Surveys Core Questions (SSCQ) 2013[1] in December 2015, SSCQ 2012 provides data for back series comparison and replaces previously publised experimental statistics. Official Statistics from SSCQ 2012 are published as online tables[2].

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[3] 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. A guide to content is provided on page **Error! Bookmark not defined.**.

The Scottish Surveys Core Questions in 2012 covered:

• self-assessed general health • disability and long-term conditions • smoking • perception of local crime rate • perceptions of police performance • highest qualification held • economic activity • household type • housing tenure	• car access • country of birth • ethnicity • religion • marital status • sexual orientation • gender • age

Further questions are under development for inclusion in future, covering income, caring responsibilities and mental well-being.

SSCQ 2014 was published in May 2016, and is also available on the SSCQ website[4].

Jamie Robertson, Sarah Martin, Michael Davidson, Julie Wilson
Surveys Branch, Strategic Analytical Policy Unit
Office of the Chief Statistician and Strategic Analysis, The Scottish Government

---

[1] SSCQ 2013: http://www.gov.scot/Topics/Statistics/About/Surveys/SSCQ/SSCQ2013

[2] SSCQ 2012: http://www.gov.scot/Topics/Statistics/About/Surveys/SSCQ/SSCQ2012

[3] About the Scottish Surveys Core Questions: http://www.gov.scot/Topics/Statistics/About/Surveys/SSCQ/

[4] SSCQ 2014: http://www.gov.scot/Topics/Statistics/About/Surveys/SSCQ/SSCQ2014

Contents

1 Source surveys and core questions

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/Topics/Statistics/16002

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 dataset with a sample size of around 20,000 responses.

Full details of the questions are available on the Scottish Government website.[5]

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 revised 2012 dataset is now published as Official Statistics as the Scottish Surveys Core Questions 2012. The website contains further information about the SSCQ project and its development.

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

The SSCQ 2012 publication was pooled from the first and second quarter year[6] of the Scottish Crime and Justice Survey 2012/13 and all four quarters each of the Scottish Health Survey 2012 and the Scottish Household Survey 2012. Responses from adults aged 16 and over were included. Due to the different sampling nature to suit the primary aims of each survey, the number of respondents vary 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 SCJS2012/13 technical report: www.gov.scot/Topics/Statistics/Browse/Crime-Justice/crime-and-justice-survey/publications/SCJS2012-13-TR

---

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

[6] This equates to half the SCJS sample and enables a consistent sample size in each SSCQ publication for 2012-15, while the SCJS is produced biennially on a financial year basis. From 2016/17 onwards, the SCJS will be produced annually, which will enable data pooled on a calendar year basis.

Briefly, the survey consists of a simple random sample, designed to achieve a minimum effective sample size of 750 interviews in the eight Police Force Areas (PFAs), enabling robust analysis at this level. One random adult per household was interviewed and asked all SSCQ and SCJS questions.

The response rate was 67.7%, resulting in 11,472 interviews during the 2012/13 financial year field work. Of these, 5,695 interviews (completed in the first two quarters of the financial year) were pooled into the SSCQ 2012 dataset. The subset was re-weighted to be in itself representative of the adult Scottish population distribution, as described for the SCJS publication.

1.2 Scottish Health Survey (SHeS) technical notes

Sampling, survey response and weighting are described in full in the SHeS 2013 technical report: www.gov.scot/Publications/2014/12/6634/downloads

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 SCJS sample is boosted with additional households to increase the number of children participating in the survey. However, these households were excluded from the SSCQ dataset: equality questions are not asked of children in the SHeS and SSCQ results cover only adult participants.

The response rate was 65.8%, and 3,804 households were interviewed in the main and health board boost samples. 4,815 resulting adult interviews were pooled into the SSCQ 2012 dataset. Of these, 3,804 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 2012 technical report: http://www.gov.scot/Topics/Statistics/16002/Methodology12

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.

67.2% of eligible households responded, leading to 10,644 household interviews. The response for the random adult interview was 62.5%, yielding 9,893 interviews. Weighting is fully described in the SHS technical report and no further weighting required for the SSCQ.

2 Weighting

Datasets from the three source surveys were combined into three new SSCQ datasets: SSCQ household variables (20,267 responses), SSCQ individual variables (20,527 responses) and SSCQ crime variables (19,516 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.

Due to collection issues in Country of Birth data in the SHS 2012 fieldwork (see section 2.1), a separate weight was required for all tabulations of the Country of Birth variable.

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[7]	5,819	5,071	3,804	2,486	10,644	8,823	20,267	16,380
Individual responses[8]	5,819	4,226	4,815	2,317	9,893	6,802	20,527	13,344
Crime responses[9]	5,819	4,065	3,804	1,526	9,893	6,583	19,516	12,174
Country of Birth AND individual responses	5,819	4,226	4,815	2,317	8,146	5,541	18,780	12,083
Country of Birth AND crime responses	5,819	4,065	3,804	1,526	8,146	5,450	17,769	11,041

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[10].

---

[7] SSCQ household variables are household type, tenure and car access
[8] SSCQ individual respondent variables are self-assessed general health, limiting long-term health conditions, smoking, highest achieved qualification, economic activity, ethnic group, religion, marital status, sexual orientation, gender and age
[9] SSCQ crime variables are perception of local crime rate and six questions on perceptions of police performance
[10] See SSCQ Weighting tables spreadsheet at http://www.gov.scot/Resource/0049/00493725.xlsx

2.1 Country of birth weights

An error in questionnaire scripting meant that in the first quarter of the Scottish Household Survey (SHS) country of birth was not collected where the highest income householder (who responds to the household survey) was also the person who responded to the random adult questionnaire. For this reason there were 1,747 cases where country of birth was not collected. Due to the nature of the error, this happens more frequently for single adult, single pensioner and single parent household types, as shown in Table 2.

This error therefore introduces a bias pattern when examining country of birth and leads to particular concerns when looking at patterns in the distribution of country of birth by variables linked with household type. For example, household type and age are closely related, with no single adult householders aged over 64 and few single parents aged over 54. Similarly there are (by definition) no single pensioners aged under 60. Any analysis of age by country of birth is biased against those individuals living alone in each age group.

To correct for this bias, the data were reweighted excluding cases where SHS failed to collect Country of Birth information on the respondent using the procedure described in section 2.

Table 2: Country of birth missing responses by household type

Household Type	Valid Response	Normal DK/Refused	SHS Missing	% cases missing	Total Sample	Weighted % missing
Single adult	2,627	5	523	16.6%	3,155	15.7%
Single pensioner	2,826	2	463	14.1%	3,291	13.3%
Single parent	861	1	121	12.3%	983	11.2%
Small adult	3,222	4	201	5.9%	3,427	7.1%
Small family	2,430	2	132	5.1%	2,564	6.3%
Older smaller	3,534	0	184	4.9%	3,718	5.5%
Large family	1,268	1	52	3.9%	1,321	3.8%
Large adult	1,993	4	71	3.4%	2,068	4.2%
Total	18,761	19	1,747	8.5%	20,527	7.6%

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.

Confidence intervals have therefore been calculated using a method to account for stratification and clustering and the resulting design effects (surveyfreq in SAS). Confidence intervals are provided alongside all survey estimates in the accompanying tables.[11] Where the intervals do not overlap then there is a significant difference between two points. If they do overlap it does not necessarily mean there is no significant difference.[12]

4 Variable list

Group	Variable name	Label
Individual	ageband	Respondent age group
	Sex	Respondent sex
	BirthCountryBand	Country of Birth
	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
	pooled_ind_wt	Adult grossing weight
	pooled_ind_wt_sc	Adult scaled weight
	pooled_ind_wt_B	Adult grossing weight for analysis of birthcountry variable
	pooled_ind_wt_B_sc	Adult scaled weight for analysis of birthcountry variable
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
	pooled_crim_wt_B	Adult grossing weight for analysis of crime and birthcountry variables
	pooled_crim_wt_B_sc	Adult scaled weight for analysis of crime and birthcountry variable
Geography	LA	Local authority

---

[11] See SSCQ Statistical tables, available at http://www.gov.scot/Resource/0049/00493724.xlsx
[12] see guidance at http://www.gov.scot/Topics/Statistics/About/Methodology/confinv

Group	Variable name	Label
	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[13] 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.

---

[13] 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 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.

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 Country of Birth

The country of birth of respondents is collected in detail by the pooled surveys. Reporting on the data in this way would uniquely identify, and therefore potentially disclose the identity of, participants from a range of countries. For this reason, country of birth has been provided as follows:
- Scotland
- England, Northern Ireland, Wales, Great Britain/United Kingdom (Not Otherwise Specified)
- Elsewhere / DK

5.2.2 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

5.2.3 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_ind_wt _B
- pooled_ind_wt_B_sc
- pooled_hh_wt
- pooled_hh_wt_sc
- pooled_crim_wt
- pooled_crim_wt_sc
- pooled_crim_wt_B
- pooled_crim_wt_B_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.