Scottish Crime & Justice Survey 2014-15 Disclosure Control Report

Introduction

This paper outlines the level of disclosure control applied to the Scottish Crime and Justice Survey (SCJS) 2014-15 Main, Victim Form and Self-completion Data Sets.

Background

With little change to the SCJS 2014-15 questionnaire and survey design the 2014-15 SCJS data sets had the same or similar level of disclosure applied to them as the 2012-13 SCJS data sets providing back to back consistent data sets for users.

The Main data set is available under an End user Licence from the UK Data Service. Both the Victim Form and Self-completion data sets are available under a Special Licence (SL) from UK Data Service due to the sensitive nature of the information recorded during these sections of the interview. In addition, all three data sets have been anonymised consistently to allow them to be merged together for more complex analysis.

Statistical Disclosure Control Methods Used

Three methods were used to disclosure control the Main, Victim Form and Self-completion data sets: re-coding, top-coding 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.

- **Top (or bottom) coding** is used on numerical variables such as age, where values of a variable above or below a specified level were grouped together and labelled as 'X+' or 'Under Z', e.g. where a respondent was aged 75 or over their age was re-coded as '75+'

- **Variable removal** is used to remove:

    o sensitive variables;

    o variables used to calculated summary variables which can be disclosive if not recoded, or used to help identify respondents in combination with other variables;

    o variables that helped with the administration of the survey fieldwork.

The following sections of this report outline the set of disclosure control approaches applied to variables in the 2014-15 SCJS Main, Victim Form and Self-completion Data Sets using each of these three methods.

Final Disclosure Control – Recoding

HBA – Health Board Areas (Main Data Set, Victim Form Data Set, Self-completion Data Set)

HBA has been recoded to align with LAA as per the change to health board boundaries implemented on the 1st of April 2014.

In addition to the boundary changes, HBAs with small base sizes have been grouped together with a mainland HBA to reduce the risk of disclosure.

SIMD_QUINT – Scottish Index of Multiple Deprivation Quintiles (Main Data Set, Victim Form Data Set, Self-completion Data Set)

SIMD quintile data zones have a number of unique Scottish data zones. As a result SIMD_QUINT

has been recoded to adsorb these unique data zones into a neighbouring quintiles within the LAA.

TABURBRUR – Urban Rural 2 fold scale (Main Data Set, Victim Form Data Set, Self-completion Data Set)

Analysing urban rural two fold data zones in combination with SIMD quintile data zones and LAA data zones reveals a number of rare Scottish data zones (< three data zones per variable combination). As a result TABURBRUR has been recoded to adsorb these data zones in to the rural category for urban respondents and vice versa for a rural respondents.

LAA – Local Authority Areas (Main Data Set, Victim Form Data Set, Self-completion Data Set)

Island LAAs with small base size have been grouped together with a Highland LAA to reduce the risk of disclosure.

QRELIG – Respondent's Religion (Main Data Set)

QRELIG is a key variable that could lead to the identification of small groups of people of different religious faiths. It was recoded to group all non-Christian religious groups together due to their small base sizes.

- Church of Scotland
- Roman Catholic
- Other Christian
- Another Religion *(Buddhist , Hindu , Muslim, Sikh , Jewish , Pagan , Another Religion)*

QDLEGS – Marital Status (Main Data Set)

- Never married AND never registered a same-sex civil partnership
- Married OR In a registered same-sex civil partnership
- Separated, but still legally married OR Separated, but still legally in a same-sex civil partnership
- Divorced OR Formerly in a same-sex civil partnership
- Widowed OR Surviving partner from a same-sex civil partnership

QDETH3 – Respondent's Ethnicity (Main Data Set)

- White – Scottish
- White – British
- White – Other *(Irish , Gypsy/Traveller , Polish , Any other white ethnic group)*
- Minority Ethnic *(Any mixed or multiple ethnic groups ; Pakistani, Pakistani Scottish or Pakistani British ; Indian, Indian Scottish or Indian British ; Bangladeshi, Bangladeshi Scottish or Bangladeshi British ; Chinese, Chinese Scottish or Chinese British ; Other Asian, Asian Scottish or Asian British ; African, African Scottish or African British ; Other African background ; Caribbean, Caribbean Scottish or Caribbean British ; Black, Black Scottish or Black British; Other Caribbean or Black background ; Arab, Arab Scottish or Arab British ; Other background)*

QBIRTH – Respondent's Birth Country (Main Data Set)
- Scotland
- England , Wales, Northern Ireland
- Republic of Ireland and Elsewhere

SOC2010 – Standard Occupational Classification 2010 (Main Data Set, Self-completion

Data Set)

- SOC2010 Major Job Groups
  - SOC2010 Sub-major Job Groups
- Manager, Directors and Senior Officials
  - Corporate Managers and Directors
  - Other Managers and Proprietors
- Professional Occupations
  - Science, Research, Engineering and Technology Professionals (309)
  - Health Professionals
  - Teaching and Educational Professionals
  - Business, Media and Public Service Professionals
- Associated Professional and technical Occupations
  - Science, Engineering and Technology Associate Professionals
  - Health and Social Care Associate Professionals
  - Protective Service Occupations
  - Culture, Media and Sports Professionals
  - Business and Public Service Associate Professionals
- Administrative and Secretarial Occupations
  - Administrative Occupation
  - Secretarial and Related Occupations
- Skilled Trades Occupations
  - Skilled Agricultural and Related Trades
  - Skilled Metal, Electrical and Electronic Trades
  - Skilled Construction and Building Trades
  - Textiles, Printing and Other Skilled Trades
- Caring, Leisure and Other Service Occupations
  - Caring Personal Service Occupations
  - Leisure, Travel and Related Personal Services Occupations
- Sales and Customer Service Occupations
  - Sales Occupations
  - Customer Service Occupations
- Process, Plant and Machine Operatives
  - Process, Plant and Machine Drivers and Operatives
  - Transport and Mobile Machine Drivers and Operatives
- Elementary Occupations
  - Elementary Trades and Related Occupations
  - Elementary Administration and Service Occupations

Responsible Person Variables (Main Data Set)
- HRP banded:
  - 1 = Respondent
  - 2 = Other household member
- HRP_AGE banded into:
  - 16-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75+

Survey Weighting Variables
Weighting variables rounded to two decimal places:
- Main Data Set:
  - WGTGINDIV, WGTGHHD, WGTGINDIV_SCALE and WGTGHHD_SCALE.
- Victim Data Set:
  - WGTGINDIV, WGTGHHD and WGTGINC_SCJS.
- Self-completion Data Set:
  - WGTGHHD_SC, WGTGHHD_SC_SCALE, WGTGINDIV_SC and WGTGINDIV_SC_SCALE

Serial Variable

The serial variable is randomised to hide information about the order the interviews were carried out in as they can be linked to local areas. This randomisation is consistent between all three data sets.

Final Disclosure Control – Top Coding

QDAGE – Respondent's Age (Main Data Set, Self-completion Data Set)
- 16 and 17 year olds bottom coded to "Under 18";
- 75+ years old respondents top coded to "75+".

NUMCAR – Number of cars in household (Main Data Set)

Variable NUMCAR was top coded to 3+ cars per household.

NOWNBIK2 – Number of bicycles in household (Main Data Set)

Variable NOWNBIKE was top coded to 5+ bicycles per household.

NUMMOT – Number of motorbikes in household (Main Data Set)

Variable NUMMOT was top coded to 2+ motorbikes per household.

NPERSONS – Number of persons in household (Main Data Set)

Variable NPERSONS was top coded to 4+ people per household.

QNADULTS – Number of adults in household (Main Data Set)

Variable QNADULTS was top coded to 3+ adults per household.

QNCHILD – Number of children in household (Main Data Set)

Variable QNCHILD was top coded to 2+ children per household.

Final Disclosure Control – Removed Variables

Household Matrix Variables (Main Data Set, Self-completion Data Set)

The following variables removed:
- QDAGE_01 to 14 (Household member's age, up to 14 household members)
- QDAGE2_01 to 14 (Household member's age banded, up to 14 household members)
- QDGEN_01 to 14 (Household member's gender, up to 14 household members)
- Person to person relationship status. (Relationship status between household members, up to 14 household members)

Interview Date Variables (Main Data Set, Victim Form Data Set, Self-completion Data Set)

The following variables were removed from the data set:
- ISSMTH (Sample issue month - 1st issue)
- INTDATE (Date of interview - Day of month)
- INTMONTH (Date of interview - Month interview conducted)
- INTYEAR (Date of interview – Year)

Sensitive Variables (Main Data Set)

Some variables are deemed too sensitive to release and are removed from the EUL data sets. The following Variables providing sensitive/ personal information about the respondent have been removed:

- QSEXORIENT (Respondents sexual orientation). This will be available by special request from the Scottish Government.
- VICFORM1 to 5 (Provides the crime code for the any victim forms completed by the respondent; these variables will be included in the Victim Form data set and the Self-completion data set we intend to put on UKDA under Special User Licence)
- PREVSEXOFF (Prevalence of sexual offences, will be covered by the Victim Form data set and the Self-completion data set under Special User Licence)
- INCSEXOFF (Number of sexual offences, will be covered by the Victim Form data set and the Self-completion data set under Special User Licence)
- LANGDIFF (Whether the respondent had any difficulty reading or understanding the questions)
- LANDINT (Whether an interpreter was used / interview conducted in another language)

Geography Variables (Main Data Set, Victim Form Data Set, Self-completion Data Set)
Variables that provide detailed geographic data that could disclose individual data zones in conjunction with other geographies have been removed. The following variables have been removed:
- URBRUR (6 folder urban/ rural classification)
- NCJBA (National Criminal Justice Board Areas)

Other Variables (Main Data Set, Self-completion Data Set)
Variables that provide non-personal data that are used to generate summary variables which could be used to disclose an individual by a user have been removed.
- Summary variable in data set: ACCTYPE – Accomodation Type.
  QDTIED, QDRENT, QACCOM, QDETACH, QFLAT, QOTH, QENTRAN, QFLOOR
- Summary variable in data set: TABNSSEC – Nat Stats Standard Economic Classification
  EC variables - QD1SEMP, QD1SUP, QD1NEMP, QD1SOWN, QD1STAF, NSSECSUB, NSSEC, NSSEC
- Summary variable in data set: ILOCLASS – Economic Activity Level
  QILO1_01, QILO1_02, QILO1_03, QILO1_04, QILO1_05, QILO1_06, QILO1_DK, QILO1_REF, QILO2, QILO3, QILO4