Scottish Crime and Justice Survey 2008/09 User Guide

Prepared by Susan McVie and Sarah MacQueen
With input from Ben Bradford and Stephanie Fohring

Scottish Centre for Crime and Justice Research
University of Edinburgh

June 2011

Contents

Introduction 1
Section 1: About the survey 2
1.1 Background to the SCJS 2
1.2 Sample Design and Methodology 5
1.2.1 Fieldwork 6
1.2.2 Questionnaire development and piloting 7
1.2.3 Structure and content of the 2008/09 SCJS questionnaire 8
1.2.4 Offence coding 17
1.2.5 Weighting 18
1.3 Methodological limitations 19
1.3.1 Sampling error 19
1.3.2 Non-response bias 20
1.3.3 Recall 20
1.3.4 Unwillingness to report 21
1.3.5 Definitions of crime 21
1.3.6 Other definitional issues 22
1.3.7 Geographic breakdown 24
Section 2: Using the data 25
2.1 Introduction 25
2.2 Access 25
2.3 File format 26
2.4 Levels of analysis 27
2.4.1 Individual level analysis 27
2.4.2 Household level analysis 27
2.4.3 Victim incident level analysis 27
2.5 Choosing a dataset 28
2.5.1 The respondent file (RF) 28
2.5.2 The victim form file (VFF) 31
2.5.3 The self-completion file (SCF) 33
2.6 Selecting the appropriate weight 34
2.6.1 Applying the weights 36
2.6.2 Calculating different weights 36
2.7 Checking and defining variables 37
2.7.1 Setting missing values 38
2.7.2 Recoding variables 39
2.7.3 Derived variables 39
Section 3: Examples of commonly used analysis 41
3.1 Examples using the respondent data file (RF) 41
3.1.1 What is the estimated number of household and personal crimes that occurred in Scotland during 2008/09? 41
3.1.2 How does experience of victimisation vary by age and sex? 43
3.1.3 To what extent does the Scottish public think that crime is a problem? And are there other issues that are considered more problematic? 50
3.1.4 What attitudes do the Scottish public have towards community sentences? 53
3.1.5 How common are credit card fraud and identity theft? 56
3.2 Examples using the victim form file (VFF) 57

3.2.1 What types of crime are most commonly reported to the police? 57
3.2.2 How does the age profile of victims of crime occurring outside pubs/bars/clubs differ from that of other victims? 60
3.3 Examples using the self-completion data file (SCF) 64
3.3.1 Is partner abuse more prevalent amongst men or women? 64
3.3.2 What percentage of the Scottish population has taken drugs, and does this vary by age and sex? 67
3.3.2 What percentage of the Scottish population has taken drugs, and does this vary by age and sex? 67
3.4 Examples that involve merging datasets 72
3.4.1 Merging variables from the RF to the VFF 72
3.4.2 Merging variables from the VFF to the RF 76
3.4.3 Merging variables from the SCF to the RF 92
Section 4: Resources for survey users 97
Appendix 98
References 100

Introduction

This User Guide is intended to provide a comprehensive introduction for anyone wishing to know more about or use the data from the 2008/09 Scottish Crime and Justice Survey (SCJS). The Guide is structured in four sections. Section 1 provides a short historical account of the development of crime surveys in Scotland, followed by technical information on the background, sample design and methodology of the 2008/09 survey in particular. This includes details of the sample design and methodology, structure and content of the questionnaire, use of weighting and some limitations of the data. Section 2 provides practical guidance on using the data, including how to access it, data file formats and levels of analysis, how to choose the appropriate dataset, selecting and using weights and defining variables. Further support for users is provided in Section 3, which uses illustrative examples to answer common analytical questions and demonstrates how to merge datasets together. The examples provided include SPSS syntax and output, which are intended to allow users to replicate analysis contained in the published reports. The final section identifies some useful resources for survey users.

The User Guide is intended to be complementary to, and read in conjunction with, the Technical Report for the survey, which was produced by the survey company responsible for conducting the fieldwork, TNS-BMRB (Page et al 2009). The data and the Technical Report are available from the Economic and Social Data Service (ESDS) Archive[1]. Links to the published reports, the ESDS Archive and other resources for survey users are provided in section four of this User Guide.

The authors of this document would like to thank Dr Catherine Millington and Stuart King from the Scottish Government's Justice Analytical Services Division for their guidance in structuring the User Guide and comments on earlier drafts. Thanks are also extended to the staff of TNS-BMRB, and Leon Page and Andrea Kinver in particular, for allowing us to draw heavily on the Technical Report and sharing their expertise and knowledge in the development of this guide.

This Guide has been prepared by a team of academics who are users of the Scottish Crime and Justice Survey and members of the Scottish Centre for Crime and Justice Research. It should be noted that the Guide is not intended to be a manual for using SPSS nor does it provide an exhaustive set of answers to each of the examples in section 3. Any opinions expressed or errors in the text are entirely our own.

[1] Previously UK Data Archive

Section 1: About the survey

1.1 Background to the SCJS

Crime surveys in Scotland have an interesting history which is not easy to piece together from existing publications. To date, there have been ten large-scale surveys to explore people's experiences of victimisation in Scotland; however, these surveys have been known under four separate names and have varied in terms of geographical coverage, survey design and methodology. The rationale behind the development of crime surveys was the rapid escalation in officially recorded crime rates following World War II, coupled with an increasing recognition of the limitations of these data in producing a 'real' measure of crime. Work commissioned by President Johnson in the US during the 1970s had shown that population surveys in which individuals were asked directly about their experiences of crime were a powerful complementary source of data to police recorded crime statistics. Moreover, crime surveys could be used to assist criminal justice planning in a number of ways, for example by finding out why people did not report crimes to the police and identifying those groups of the population at most risk of being victimised.

Following exploratory work in England in the late 1970s, a group of academics and researchers from the Home Office Research and Planning Unit devised proposals for a national survey of crime in England and Wales. These proposals were agreed by the Home Office in 1981 and, shortly thereafter, plans to extend the survey to Scotland were agreed by the Home and Health Department of the, then, Scottish Office. According to Anderson (1999), there were four main factors behind the eventual launch of the British Crime Survey (BCS) in the early 1980s:

> "a build up of pressure from the academic community; the feeling that Britain was lagging behind other countries; the influence of so-called 'administrative criminology', focused on questions of crime prevention and efficient resource allocation, within the Home Office Research and Planning Unit; and the prevailing climate of 'law and order' politics, heightened by the riots in many English inner cities in 1981." (p48)

The BCS was launched in 1982, covering England, Wales and central and southern Scotland. The main findings from the BCS Scotland, as it was known, were published by Chambers and Tombs in 1984. The BCS was repeated again in England and Wales in 1984, but Scotland did not participate mainly for reasons of cost, but also because it was felt there had been insufficient time to consider the results of the first survey. Scotland participated again in the 1988 British Crime Survey, but there were a number of problems with the English-based survey company, which included delays in the completion of the fieldwork and errors in the calculation of the victimisation rates. As a result of these problems, the main findings from the Scottish component of the BCS were not published until four years later (Payne 1992).

A decision was taken in the early 1990s to create a new, independent Scottish Crime Survey (SCS), mainly because it was felt that the Scottish Office would be able to exert more control over the survey company if they were directly responsible for commissioning the fieldwork. In addition, there was a desire within the Scottish Office to change certain aspects of the survey design, such as its geographical coverage. The BCS Scotland surveys had been restricted to central and southern Scotland, with no interviews conducted north of the Caledonian Canal. Therefore, when the new SCS was launched in 1993 it was extended to include the experiences of those living in Grampian, the Highlands and the largest of the Scottish Islands, with no reduction to the sample size in central and southern Scotland (Anderson and Leitch 1996). Tweaks were also made to the survey questionnaire to bring it more in line with the interests of policy makers north of the Border. On the whole, however, the structure and content remained similar enough for comparisons to be retained with the BCS in England and Wales. A further three sweeps of the SCS were repeated in 1996 (MVA 1998), 2000 (MVA 2002) and 2003 (McVie et al 2004).

The Scottish Crime Survey was carried out rather sporadically during these early years and the design and sample size remained essentially unchanged. The British Crime Survey, on the other hand, underwent a number of changes and improvements. From 1992 to 2000, the BCS was carried out on a biennial cycle. The methodology for the survey switched from old fashioned paper and pencil interviewing (PAPI) to computer assisted personal interviewing (CAPI) in 1994. Then in 2001, the Home Office moved to a continuous survey design with an increased sample size of 40,000, gathering data from around 3,300 respondents every month. In comparison to the SCS over the same period, the BCS was being well utilised by its customers and had been established as a central tool for informing policy makers, practitioners and academic debate.

A fundamental review of the SCS was commissioned by the Scottish Executive in 2002. This coincided with a period of both managerial and institutional transformation during the early years following devolution, as well as a review of the role of victims as stakeholders in the criminal justice system (see McAra 2008). The aim of the review was to conduct a thorough appraisal of the design, content and management of the existing survey and consult widely to clarify the main objectives of the survey and identify an appropriate design strategy for the future. In particular, the Scottish Executive wished to identify the most cost-effective method of increasing the sample size to allow more accurate reporting of crime rates at both national and local levels. The review, conducted by independent researchers who had previously worked as contractors on the survey, made three key recommendations: the sample size should be significantly increased; the data collection method should become computerised; and the survey should be continuous (McCaig and Leven 2003).

In 2004, the Scottish Executive re-launched the survey again as the Scottish Crime and Victimisation Survey (SCVS) using a computer assisted telephone interviewing (CATI) methodology with an intended annual sample size of 27,500. The option of a face-to-face survey with a sample size of 16,000 respondents was considered, as it would have cost approximately the same. However, the desire

for a large sample to be able to present data at smaller areas of geography was great. McCaig and Leven had argued strongly in favour of the telephone survey and, although outlining some of the risks, voiced the opinion that:

> "Our expectation is that the move to telephone will not lead to a variation in victimisation recording rates that would be significant enough to impact on policy studies to a worrying extent." (McCaig & Leven 2003: 60)

While confidence was high that the telephone survey would be a success, a calibration exercise involving a scaled-down face to face survey of 3,000 respondents was commissioned to run in parallel to the main survey in order to provide a measure of victimisation against which the telephone survey could be compared. The calibration exercise was carried out by a separate survey company to that conducting the telephone survey, and proved to be vital. Major problems were experienced with the telephone survey and the sample achieved was less than 10,000. A detailed report by Hope (2005) presents the findings of the comparison exercise and highlights the various problems of bias encountered with the telephone survey. Hope concludes:

> "…we have not found sufficient evidence to conclude that the telephone survey is likely to be accurately measuring victimisation. We have been unable to devise a weighting approach that satisfactorily corrects the many demographic biases that are observable in the data but even if these biases could be corrected, our conclusion is that the telephone survey suffers a more fundamental and irresolvable bias resulting from what appears to be a tendency for non-victims to be more likely to refuse to participate." (Hope 2005: 41)

As the robustness of the data for the 2004 SCVS could not be substantiated it was not feasible to produce victimisation rates from the telephone survey. Instead, the results from the much smaller face to face survey were published (Hope 2007). On the advice of the SCVS Technical Advisory Group, which includes independent academic and research advisors, the telephone survey was discontinued. Scottish Government reverted back to a face to face methodology for the 2006 Scottish Crime and Victimisation Survey, with a sample of around 5,000 (the same as that for the four sweeps of the SCS), although there was one improvement in that CAPI was used to collect the data instead of paper based questionnaires (Brown and Bolling 2007). A guide to using the 2006 SCVS data is also available (Norris and McVie 2009).

Following the 2006 SCVS, a further review of the crime survey was carried out by analysts within the Scottish Government Justice Department, which included consultation with both internal policy colleagues and external users, via the CJ-Quest Network of the Scottish Centre for Crime and Justice Research (McVie and Norris 2007). The general consensus was that a continuous survey with as large a sample size as possible was still desirable, but face to face interviews were essential, as per the BCS model. In particular, it was noted that the survey should allow results to be published at Police Force Area level. As a result, the new Scottish Crime and Justice Survey (SCJS) was launched in April 2008.

The subject of this User Guide, the SCJS 2008/09 is the ninth national survey of victimisation in Scotland, and represents the first of another new series of crime surveys which differs substantially from its predecessors. The SCJS is a continuous survey, which means that the fieldwork was spread out over a 12 month period, from April 2008 to March 2009. The shift to cover financial years is in line with other national statistics produced by Scottish Government, including police recorded crime data. The survey involves face to face interviews, a sample size of 16,000 and computerised methods are used to collect the data. Further details of the sample design and methodology are provided below, and can be found in the Technical Report for the survey (Page et al 2009). The findings from this survey were published in MacLeod et al (2009), who make the point that "care should be taken if comparing the results from the SCJS 2008/09 with sweeps from previous Scottish crime surveys" (p1).

The 2009/10 SCJS, and the tenth national survey in Scotland, has recently been published (Page et al 2010a), and the 2010/11 survey is in the field at the time of writing, although the sample for this sweep has been reduced to 13,000. While the intention at the time of commissioning this latest series of surveys was to maintain an ongoing continuous survey indefinitely, this decision was subsequently overturned due to the strain on resources within the Scottish Government as a result of the global economic downturn. A further sweep of the SCJS is scheduled to take place in 2012/13, again with a reduced sample size of 13,000. It remains to be seen what other changes may affect this survey.

1.2 Sample Design and Methodology

The Scottish Crime and Justice Survey 2008/09 involved face to face interviews with a sample of 16,003 adults aged 16 or over who were living in private residential addresses in Scotland during the period between April 2008 and March 2009. Fieldwork was undertaken continuously over this period, with approximately 1,333 interviews conducted each calendar month. The survey was conducted face-to-face in the respondents' homes and was administered by specially trained professional interviewers using Computer Assisted Personal Interviewing (CAPI). The survey also included a short self-completion section on sensitive topics which respondents answered using Computer Assisted Self-Interviewing (CASI).

The 'survey universe' of the SCJS was all households living in private residential accommodation in Scotland at the time that the sample was drawn and, within those households, all individuals aged 16 or over. The survey universe provides the population from which the sample is drawn which, in this case, includes the whole of mainland Scotland and all but the smallest inhabited islands. The sample frame for SCJS 2008/09 was the Small User file of the Postcode Address File (PAF), which was expanded using the Multiple Occupancy Indicator (MOI). The sample frame provides the list from which addresses are selected for contacting. Once a household was selected for interview an attempt to contact was made and, where contact was made, details of household members collected. One adult from that household was randomly selected for interview.

The SCJS 2008/09 sample design was different from those of the preceding SCVS and SCS surveys in a number of respects. Most importantly, its planned annual sample size of 16,000 interviews was considerably larger than previous surveys (5,000 for most surveys); and the survey design required the equivalent of at least 1,000 simple random sample interviews in each Police Force Area (PFA). A further stipulation of the survey design was that interviews be spread out evenly across the year. Finally, while the previous surveys had completely clustered designs, the majority of the SCJS sample was un-clustered; clustering only occurred in the more sparsely populated 'rural' areas of Scotland.

The requirement for 1,000 interviews per PFA meant that a disproportional survey had to be implemented whereby the PFAs with the smaller populations were over-sampled relative to their populations and those with the largest populations under-sampled (their samples were reduced to maintain the total sample size at 16,000 interviews). This has the effect that people living in rural areas of Scotland are over-represented in the SCJS, while those living in urban areas are under-represented. While weighting (see below) will take care of this in statistical terms, it may be that the over-representation of people living in low-crime rural areas will have affected the types and raw numbers of crimes covered by the survey at the national level. This is a reasonable assumption given the design change, as we know that crime is not evenly distributed, although it has not been objectively tested.

1.2.1 Fieldwork

Fieldwork for the SCJS 2008/09 took place between the 1st of April 2008 and the 31st of March 2009. The sample was more or less evenly distributed throughout the 12 month fieldwork period, although there was some degree of 'front-loading' of the interviews in order to ensure that the fieldwork would be completed within the desired timescales. The fieldwork was carried out by specially trained interviewers and supervised by experienced managers who conducted thorough checking throughout the fieldwork period. For example, 12% of addresses where a successful interview was obtained were re-contacted ('back-checked') to verify that the interviewer had conducted the interview and that key details they had collected were correct.

All households selected for interview were sent a letter from the Scottish Government in advance of an interviewer calling at the home address. The letter explained a little about the survey, why the address had been selected and informed the occupiers that an interviewer from the survey company would call within the next few days. Included with the advance letter was a leaflet from the Scottish Government which provided people with further details about the survey, including some example findings from previous surveys. The leaflet also tried to answer some questions that potential respondents might have, such as issues relating to confidentiality. No incentives were offered to potential respondents and participation in the survey was entirely voluntary. Importantly, the respondent for each household was selected at random, and no substitution was permitted if that individual chose not to participate.

The survey was administered in two distinct parts. The first part (or main survey) was a Computer Assisted Personal Interview (CAPI) that involved the selected member of the household being led through a pre-programmed questionnaire by the interviewer and all responses being entered into a tablet PC. The final adjusted response rate for the 2008/09 main survey was 71% (n=16,003).

The second part of the survey involved a self-completion questionnaire covering illicit drug use, stalking and harassment, partner abuse and sexual victimisation. Due to the sensitive nature of the questions, respondents were encouraged to enter their answers directly into the tablet PC used by the interviewers, a method known as Computer Assisted Self Interviewing (CASI). Although all respondents were encouraged to use the computer themselves, interviewers were allowed to administer the questionnaire for those who preferred this approach. Respondents were shown the screen and the interviewer tapped the appropriate answer. A total of 69% of respondents to the main survey answered the self-completion questionnaire. Lack of time was the most common reason given for not-completing this part of the questionnaire.

The proportion of respondents that filled in the self-completion questionnaire varied across different groups. For example, while equal proportions of males and females answered the self-completion section (69% of all men and women who completed the main survey), participation in the self-completion section decreased with age (so 76% of 16-24 year olds answered the self-completion section compared with 61% of those aged 60 or over). Similarly, a lower proportion of those living in the 15% most deprived areas answered the self-completion section compared with those living in the rest of Scotland (58% and 71% respectively).

1.2.2 Questionnaire development and piloting

Because the 2008/09 SCJS represented a new model of victimisation survey in Scotland compared to those that had come before, the questionnaire underwent an extensive period of development and piloting. In the first instance, the Scottish Government undertook consultation internally with policy colleagues with an interest in the survey, and the CJ-Quest network of the Scottish Centre for Crime and Justice Research (SCCJR) conducted an external review of the 2006 SCVS questionnaire (McVie & Norris 2007). The results of both of these fed into the first draft questionnaire for the SCJS 2008/09.

Following these reviews, the questionnaire was developed jointly by the Scottish Government and the survey company. Once agreed, extensive piloting was carried out before interviewing started for the 2008/09 survey in April 2008. The piloting involved a number of stages:

- Cognitive question testing, designed to test respondents' understanding of specific questions in the questionnaire identified as potentially ambiguous, difficult to answer, or in some other way problematic during the questionnaire development phase.

- Consultation with specialist groups. The draft self-completion questionnaire sections covering partner abuse and sexual victimisation were discussed with representatives of three organisations concerned with the domestic abuse and sexual victimisation of males and females – Rape Crisis, Scottish Women's Aid and the Open Road Project.
- CAPI pilots to test the computerised questionnaire programme.

1.2.3 Structure and content of the 2008/09 SCJS questionnaire

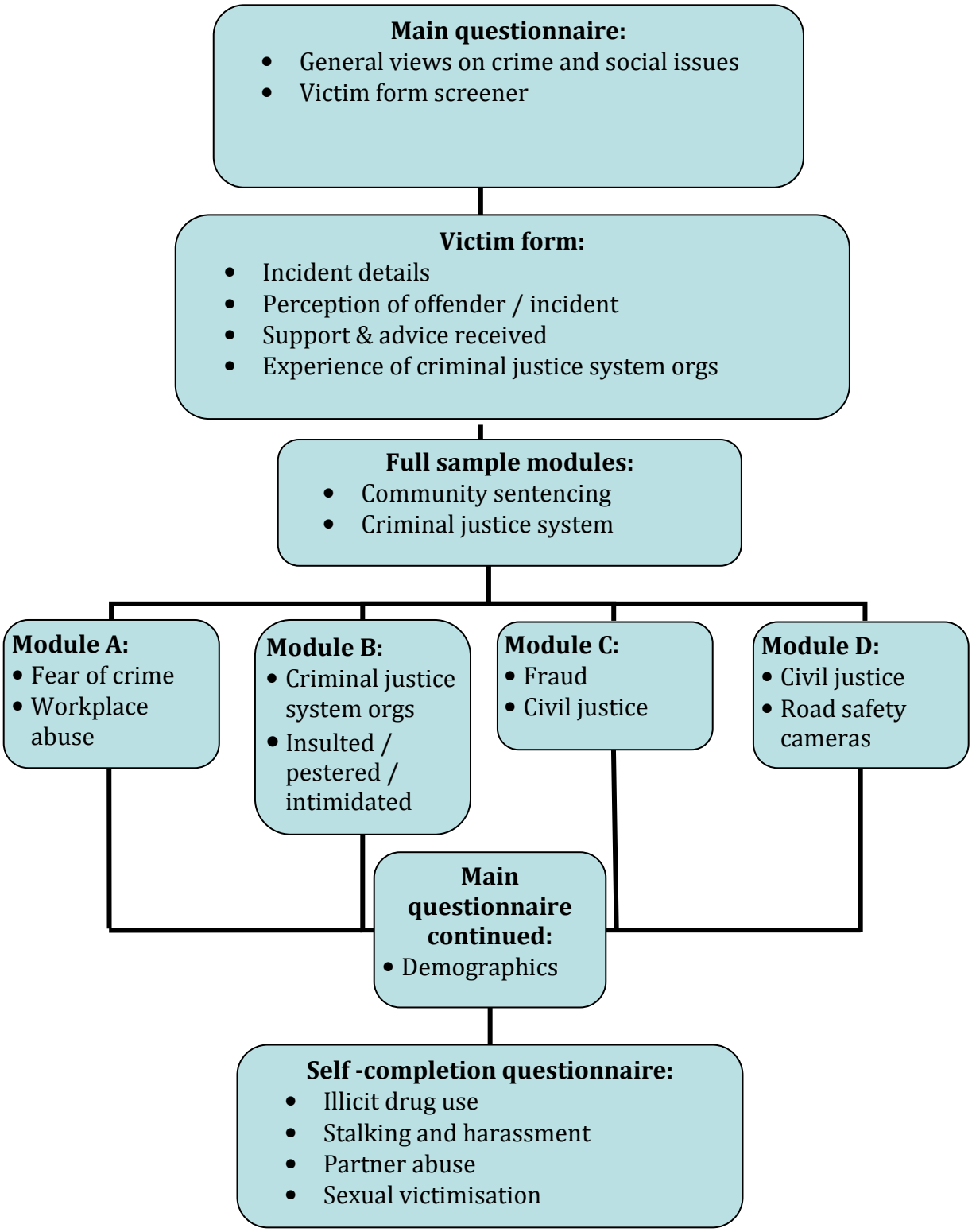
The SCJS questionnaire consists of different sections and modules that are designed to collect a wide variety of information about the respondents' experiences, knowledge and attitudes. The complex structure comprises three core elements (Page et al 2009):

- The main questionnaire, which includes: screener questions to elicit types of victimisation experienced; a set of core modules asked of the whole sample; and a set of quarter-sample modules, containing questions on a variety of topics;
- A victim form which collects details about incidents of victimisation a respondent may have experienced (respondents may be asked to complete up to 5 of these);
- A self-completion questionnaire covering sensitive issues (respondents can refuse to answer this section if they do not wish to complete it).

Each of these elements contains different sections and are inter-related (for example, there are checks in the self-completion questionnaire to make sure an incident reported in the victim form does not need to be reported again in the self-completion questionnaire). Within some sections there is further filtering so that questions are only asked of small sub-samples (for example, those who have had contact with the police in the last year). Such question 'routing' in the SCJS and similar questionnaires can sometimes be quite complex, and before starting analysis it is advisable that users acquaint themselves with both the general structure of the questionnaire and the specific questions they wish to use. The full questionnaire is available from the ESDS Archive and on the survey website – again, users are strongly advised to download and consult it before commencing analysis.

Figure 1.1, below, provides an overview of the questionnaire structure, including the various component parts of the survey instrument. Further, more detailed explanation of the questionnaire content is provided in the text and boxes that follow.

Figure 1.1: Questionnaire Structure of the 2008/09 Scottish Crime and Justice Survey



Source: 2008/09 SCJS: Technical Report (Page et al 2009 : 24)

Main questionnaire (1)

The first part of the main questionnaire contains two sections. The first covers attitudes and opinions surrounding crime issues, while the second comprises the victim form screener (see Box 1).

Box 1: Main questionnaire (1) *(16,003 respondents)*

1. General views on crime and social issues: attitudinal questions on how important various social issues, including crime, are in Scotland; questions about the local area, including how long the respondent has lived in the local area, how much the crime rate has changed, and how safe the respondent feels; how worried respondents are that specific crimes will happen to them and opinions on how likely they are to be a victim of these crimes. The majority of this section is administered to all respondents.

2. Victim form screener. Respondents are asked whether they have experienced certain types of crime in the reference period. The screener crime types are separated into three broad groups: vehicle incidents; household property incidents; and personal incidents. The screener questions are designed to identify if the respondent experienced any incidents that fall within the scope of the SCJS (see 'Victim Form' below). Responses to the screener questions trigger the use of the victim form questionnaire if a respondent has experienced at least one such incident.

Victim form

Each respondent who reports experiencing at least one incident of crime at the screener stage will complete a victim form. The victim form questionnaire is designed to elicit all of the details of an incident, irrespective of what type of incident the victim form was triggered by. At the screener stage respondents are asked to report all incidents that they experienced in the reference period. A maximum of 5 victim forms are completed per victim (regardless of the number of incidents the respondent reports) so as not to overburden the respondent. Incidents are covered in a specific priority order that broadly reflects the rarity and seriousness of certain crimes. Generally speaking, all personal incidents are asked before property-related incidents, which are asked before vehicle-related incidents. Where the respondent has experienced more than one incident of the same type of crime, the most recent incident is asked about first. However, where the respondent declares that these multiple incidents are part of a 'series' (defined as 'the same thing, done under the same circumstances and probably by the same people' in the survey questionnaire) only one victim form is completed. Details of the most recent incident are collected and it is assumed that all the other incidents in the series had the same characteristics. Again, this is mainly done to reduce respondent burden.

In the SCJS 2008/09, 23% of all respondents reported at least one incident of victimisation: 15% of respondents completed a single victim form only, while only 1% completed five victim forms (the maximum allowed). In total 5,772 victim forms were completed by 3,619 respondents. The victim form comprises three main sections, as shown in Box 2, below.

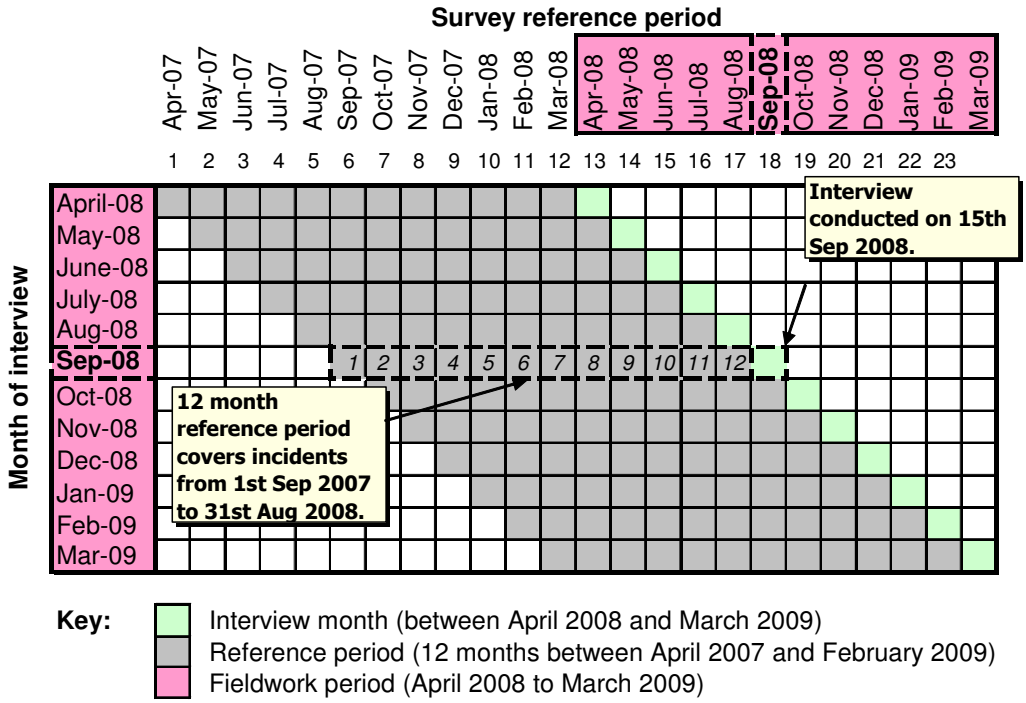
Box 2: Victim Form *(5,772 completed forms)*

1. Specific details about the incident or series of incidents, including:
   - the exact month(s) in which the incident(s) occurred (or a specific quarter if the respondent is unsure);
   - the respondent's description of the incident; and
   - important characteristics of the incident (for example, where the incident took place, whether anything was stolen or damaged, whether injuries were caused, etc.)
2. Perception of the offender and the incident. Questions include:
   - whether anyone other than the offender was responsible for the incident; whether the respondent themselves used force or were under the influence of drugs or alcohol prior to the incident;
   - details of the prosecution of the offender (if applicable); and
   - opinions on what sentence the offender should have received, if any, and whether the respondent considered the incident to be a crime or not.
3. Support and advice received. These questions pertain to the aftermath of the incident, and include:
   - what advice and support the respondent would like to have received from various organisations, what they did receive and from whom, and how satisfied they were with it;
   - whether the police came to know about the matter, how they came to know and how satisfied they were with the police handling of the incident;
   - information and assistance received from the police, Victim Support Scotland and the Witness Service, and what the respondent would like to have received.

The main reason for introducing crime surveys to the UK was to provide a measure of victimisation that is as comparable as possible to the police recorded crime statistics. For that reason, the scope of the SCJS only includes incidents that happened within Scotland and within the previous 12 month reference period. For incidents occurring on-line, if the respondent was living in Scotland at the time of the incident, then the incident would be included. In cases where the incident occurred outside of Scotland or outwith the reference period then the victim form questionnaire terminates (unless the incident occurred within the month of interview, in which case the data is collected but not included in the survey crime count) and the questionnaire moves on to the next victim form or the start of the next part of the main questionnaire. Note that because the SCJS is

conducted over the course of a financial year, with interviews conducted on a monthly basis, the total reference period for the survey is actually 23 months, as illustrated by Figure 1.2.

Figure 1.2: Reference period structure for the 2008/09 SCJS



Source: Adapted from the 2008/09 SCJS: Technical Report (Page et al 2009: 54)

Main Questionnaire (2)

The second part of the main questionnaire consists of two full sample modules (i.e. asked of all respondents), which cover different issues relating to community sentencing and the Scottish criminal justice system. Box 3 describes the areas that these two modules cover.

Box 3: Full Sample Modules *(16,003 respondents)*

1. Community sentencing. Respondents are asked about what community sentences they are aware of, how they became aware of them, whether the sentences would make an offender less likely to commit a crime in the future, and whether they agree or disagree with a series of statements about community sentences and prisons.

2. Scottish criminal justice system. Questions here cover respondents' level of awareness of the system as a whole and confidence in it; which of the component organisations that make up the criminal justice system they have heard of, and which they have personally been in touch with; and respondents' confidence in their local police force.

Quarter-sample modules

A further set of four modules are included in the questionnaire covering a wider range of topics. Addresses are randomly allocated to one of the four modules at the sampling stage, so in theory one quarter of the addresses are allocated to each and they are evenly distributed (although, in practice, differences in response rates lead to slightly uneven distribution in the final dataset). Boxes 4 to 7, below, describe each of the four modules.

Box 4: Module A *(4,027 respondents)*

- Fear of crime. Includes questions about feeling safe when walking alone after dark, why respondents feel unsafe, how common they think various crimes are and where they have got this impression from. All of these are with reference to the respondent's local area (within a 15 minute walk of the respondent's home).
- Workplace violence (asked only of respondents who are currently in employment and deal with members of the public as part of their job, either face-to-face or on the telephone). Questions cover verbal or physical abuse at work, how often it happens, whether it was reported to the employer, whether they have received training to deal with aggressive behaviour and the extent to which they worry that abuse at work affects their health.

Box 5: Module B *(4,004 respondents)*

- Criminal justice system organisations. Questions cover:
  - respondents' experience of contacting the police in the reference period, including how the contact was made and how many times, why the contact was made, satisfaction with the police's handling of the matter and, where applicable, reasons for dissatisfaction;
  - the same set of questions about contacts initiated by the police;
  - opinions of whether the police in their local area do a good job; and
  - questions on awareness, contact and satisfaction with contact with regard to the Procurator Fiscal.
- Forms of harassment. Respondents are asked if anyone who is not a member or their household has insulted, pestered or intimidated them in any way in the 12 month reference period, either in person or by some other means, and if so how many times it has happened. Follow-up questions include:
  - by what means it happened;
  - how many people were involved;
  - how well, if at all, the respondent knew the perpetrators;
  - why it happened; and
  - whether it may have been racially or religiously motivated or related to sectarianism, and if so how.

Box 6: Module C *(3,980 respondents)*

- Card fraud and identity theft. This section focuses on forms of fraud that are not covered by the victim form screener section of the main questionnaire. Where incidents of this type are recorded in a victim form, some details are transferred to this section to save repeating questions.
    - Card fraud covers the unauthorised use of credit and bank cards to buy or pay for things or withdraw cash, and the use of card details for the same purposes.
    - Identity fraud relates to someone pretending to be the respondent or using their personal details to do things such as obtain credit, open a bank account, apply for a mobile phone contract or state benefits or for official documents such as a driving licence or passport or to commit some other kind of fraud.
- Civil justice. This section relates to four types of problem or dispute that may be dealt with under civil justice procedures, including: home, family or living arrangements (e.g. neighbours, family, housing and immigration); health and well-being (e.g. accidental injury, medical negligence or mental health issues); money, finances or any goods/services paid for (debt, benefits and faulty goods and services); unfair treatment (e.g. discrimination, employment related issues). Follow-up questions include:
    - whether it was resolved or not;
    - if they have solved or are trying or planning to solve the problem they are asked if they are using help and advice from others;
    - if the problem has been solved then they are asked how satisfied they are with the results.

Box 7: Module D *(3,991 respondents)*

- Civil justice. The bank of questions from Module C is repeated in Module D to achieve responses from half the sample.
- Road safety cameras. Road safety cameras are defined as both speed cameras and red traffic light safety cameras. Respondents are asked how far they agree with the use of each, and whether they agree or disagree with a battery of statements about them.

Main questionnaire (3)

The final part of the main questionnaire collects a range of demographic information about both the characteristics of the individual respondent and the household. The information collected at this stage of the interview is detailed in Box 8.

Box 8: Main questionnaire (3) *(16,003 respondents)*

- Age
- Gender
- Marital status
- Ethnicity
- Religious background
- Tenure and property type
- Newspaper readership
- Health status
- Employment status, including questions to allow Office for National Statistics Socio-Economic Classification (NS- SEC) coding
- Household income and ability to afford an unexpected expense
- Age and gender of other persons in the household are collected at the interview screening stage.

Self-completion questionnaire

Finally, all respondents are asked if they would be willing to complete a further questionnaire themselves by entering their answers directly into the tablet PC. Not all respondents to the main part of the survey choose to complete the self-completion questionnaire (in 2008/09 69% of respondents to the main survey also completed this separate questionnaire). The self-completion survey contains four sections, as shown in Box 9, each of which covers sensitive topics.

Box 9 : Self-completion questionnaire *(10,974 respondents)*

1. Illicit drug use. Respondents are asked whether they have ever used any of 16 types of illicit drugs. A series of follow-up questions are asked of drug users:
   - drug use in the last 12 months and last month;
   - types of drugs used most and difficulty of obtaining them;
   - nature of first drug use, including age, type of drug and methods of drug taking;
   - experience of multiple drug use or drug use combined with alcohol;
   - level of dependence on drugs and efforts to reduce drug use.

(continued overleaf)

2. Stalking and harassment, partner abuse and sexual victimisation

This section begins with a screener section collecting information about respondents' relationship history since age 16 as the questions about partner abuse are only asked of respondents who have had a partner since the age of 16. Respondents are also asked about their sexual orientation in the self-completion section, as it was felt that the self-completion format would encourage a higher response rate. For each of the sets of questions below, where respondents report more than one incident, details are collected for the most recent one only.

2a. Stalking and harassment. Respondents are asked whether they have experienced any of a series of behaviours that would be defined as stalking or harassment during the last 12 months. Those that have are asked further questions about the nature of their relationship to the perpetrator. The respondent is also asked whether the police came to know about the incident, and if not, why not.

2b. Partner abuse. This section is only asked of those who have had a partner since the age of 16. Respondents are presented with different forms of psychological and physical abuse and then asked if they have experienced any of these since age 16, and if so, how many partners perpetrated these acts. If abuse has occurred within the 12 month reference period, a series of follow-up questions are asked. All those who have had a partner since they were age 16 are asked whether they consider themselves to have ever been a victim of domestic abuse. The term is not defined to the respondent.

2c. Sexual victimisation. This section covers certain types of sexual offences, which are classified as 'serious sexual assault' and 'less serious sexual assault'.[2] Serious sexual assault includes: actual or attempted forced sex; forced or attempted other sexual activity. Less serious sexual assault includes: indecent exposure; sexual threats; and unwanted sexual touching. Victims of sexual assault are asked follow up questions about the perpetrator, location, timing and frequency of incidents, as well as whether the police were informed. Victims of *serious* sexual assault are also asked about the injuries sustained, whether the case resulted in prosecution or conviction, who came to know about it and whether drugs/alcohol were involved.

[2] The terms 'less serious sexual assault' and 'serious sexual assault' are adopted to distinguish between the two types of sexual assault which were asked about separately in the questionnaire. This is consistent with the practice adopted by the Home Office in reporting of the BCS. The terms do not relate to the seriousness of the impact on the individual experiencing an incident, as this may vary according to the particular circumstances of an incident.

1.2.4 Offence coding

Correct coding of the offences reported to the SCJS is essential if SCJS data are to be comparable with police recorded crime data. This is a difficult process, and full comparability is extremely difficult to achieve due to the uncertain nature about how crimes would actually have been dealt with by the police.

Until recently, the offence coding system used for crime surveys in Scotland was based on that developed for the 1982 BCS; however, just prior to the 2008/09 survey this was revised to more accurately reflect the Scottish system of justice and, in particular, changes to the police crime reporting standards introduced in 2004. The coding procedures are designed to match as closely as possible the way incidents would be classified by the police to aid comparison. This involves collecting detailed information about incidents in the victim form section of the questionnaire. All victim forms are reviewed by trained coders who refer to the SCJS Offence Coding Manual in order to determine whether what has been reported in the interview represents a crime or not and, if so, what offence code should be assigned to the crime based on the most likely way in which the police would have recorded the offence.

The offence coding system consists of a number of steps, carried out by both the survey company and the Scottish Government. This involves multiple checks of 'hard to code' offences and a quality control check of 10 % of all certain codes. The end result of the process is that every victim form has a final offence code assigned to it, which is agreed following discussion between the Scottish Government and the survey company. The dataset also includes variables that show any codes assigned at the intermediate stages (for example, as offences are recoded from one category to another the original coding is kept as a check on the history of the process).

The available SCJS data files include all the offence codes that have been assigned to each victim form at each stage of the coding process. The variables in the victim form data file which detail this are:

- VOFFENCE: code assigned by the original coder;
- SOFFENCE: code assigned by the supervisor;
- FINLOFFC: code assigned by the Scottish Government research team;
- OFFENCE: final offence code assigned.

Once offences have been coded they are generally aggregated up into crime groups for analysis. The main reason for this is that many of the 63 specific offences collected in the survey are found in very small numbers. Therefore these need to be grouped together in some sensible way for analysis purposes. Details of how the offence codes are grouped together are shown in tables A1 (crime groups) and A2 (aggregate crime groups) which can be found in the Appendix to this report.

1.2.5 Weighting

Weights are calculated for the SCJS sample for a number of reasons. These include:

- Correction of the sample for unequal probabilities of selection that arise from various aspects of the sample design. For example, the requirement for a final sample in each PFA equivalent to a simple random sample of 1,000 means that PFAs with smaller populations are over-sampled relative to more populated PFAs. Correction weights such as these are known as design weights.
- Differing response rates by sub-groups within the sample. Response rates can differ by household type, age, and gender (for example, a young adult male living alone may be less likely to respond to the survey than one living with a partner and child). These are known as non-response corrections or calibration weighting.
- To enable results from the survey to be reported in terms of the population of Scotland. A weight that includes an expansion factor is required to gross up the sample data to allow the results to be expressed as population values.[3]

Types of weight

The SCJS provides three different units of measurement: households, individuals and crime incidents. At various points in the survey the adult being interviewed provides information on behalf of the household as a whole, while at other times they are speaking about their own experiences, attitudes or opinions. Those who have experienced victimisation also provide information about incidents of victimisations they or their households as a whole have suffered in the previous year (the 'reference period' – see figure 1.2, above). An individual and/or household can have more than one victimisation reported in the survey, whereas other pieces of data about the household or individual are collected only once. There are, therefore, three main units of analysis used in the SCJS: households, individuals and crime incidents. The weight used is dependent on the unit of analysis and the dataset being analysed (weighting is covered in more detail in section 2.6).

(1) Household weights are constructed for use when the household is the main unit of analysis. Some crimes are considered household crimes (e.g. housebreaking, vandalism to household property, theft of or from a motor vehicle) and therefore the main unit of analysis is the household. Similarly, analysis for certain demographic questions that apply to the whole household requires a household level weight (for example, accommodation type).

(2) Other crimes are considered personal crimes (for example assault or robbery). Here the main unit of analysis is the individual and the individual

[3] An expansion factor scales all cases up by the same amount (i.e. it is equally applied to all respondents); whereas, a weight is differential (i.e. each case is weighted by its characteristics).

weight applies. The individual weight should also be used when analysing individual or attitudinal data, such as personal feelings of safety when walking alone in the local area and other questions where the respondent is asked for their personal opinion or information about themselves.

(3) Incident weights are comprised of the household and individual weights, described above, with an additional component which reflects whether incidents in the victim form reflect a single or a series incident. The incident weights should be used for all analysis conducted on the victim form file (VFF) if crime variables are being analysed.

Data users should note that separate household and individual weights are created for the self-completion questionnaire. These weights are constructed on the same basis as the main individual and household weights, but are based only on the sample of respondents who answered the self-completion section of the questionnaire.

1.3 Methodological limitations

Like all sample surveys, the SCJS has a number of limitations that users should bear in mind. Many of these stem from one source: the fact that the survey is administered to only a small proportion (or sample) of the population it is attempting to cover (here, adults in Scotland who live in private households). Sample surveys are designed to provide estimates of population values – the SCJS provides an estimate of the number of people who are victims of crime in Scotland, for example. The values estimated by the survey will differ from the true population values for a number of reasons. These reasons can be grouped into two main types of problem: bias, which will lead to systematic differences between survey estimates and population values (for example because of errors in the design of the survey); and variance, which arises from the nature of the sampling techniques used and which does not lead to systematic differences between estimates and population values.

Variance can often be accounted for by various methodological and statistical tools, many of which are built into SCJS design and the datasets made publicly available. Bias can be more difficult to deal with, and some level of bias is almost inevitable in any sample survey, no matter how well designed.[4] Some of the main limitations caused by variance and, particularly, bias are outlined below.

1.3.1 Sampling error

Sampling error is a fundamental limitation of sample surveys. Unless the entire population is interviewed (a census) the estimates produced from such surveys will always differ from the population values they represent to a greater or lesser degree.

[4] See sections 2, 3, 4, and 5 of the 2008/09 SCJS: Technical Report (Page et al 2009) for detail on the steps taken in the design of the SCJS to reduce the impact of these issues.

In general, the larger the sample the smaller the sampling error will be, since larger samples will produce estimates with smaller variance (smaller standard errors). Although the SCJS is a very large survey relative to the population it covers, because of the highly unusual nature of some of the crimes and experiences covered, small sample size can still be an issue (for example in analyses looking at victims of rare crimes, such as robbery). In such cases sampling errors caused by small sample sizes can be a real problem, because the survey estimates produced have large margins of error.

1.3.2 Non-response bias

While the 2008/09 SCJS achieved a 71% response rate to the main survey (MacLeod et al 2009), 'non-response' can introduce a significant amount of bias into the sample. This is a particular problem for crime surveys as those population groups who are least likely to agree to complete the survey are also those disproportionately likely to be victims of crime – in particular young men (see for example Lynn 1996, cited by Lynn and Elliot 2000). If groups that are particularly likely to be victims of crime have systematically higher levels of non-response this could undermine the accuracy of the victimisation rate and other headline SCJS indicators.

There are some ways to address non-response bias, such as population weighting (where individuals from under-represented groups will be given higher weights than those from over-represented groups). However, it is not possible to rely entirely on weighting to adjust for missing respondents. For example, young adult males are typically under-represented in social surveys. Weighting can be used to bring the sample in line with expected population proportions; however, the young adult males in the sample may not be typical of those who choose not to respond in terms of their other characteristics or experiences. In addition, non-response can cause problems in terms of estimating rare crimes. Where very small numbers of cases – for example people who experience robbery – are weighted up to give population estimates, the presence or absence of even a few cases can have a major effect on the accuracy of the estimates. Users should be cautious about calculating estimates for individual offence codes with small numbers of respondents due to the wide confidence interval that are attached to such incidents.

1.3.3 Recall

The SCJS asks respondents about incidents that occurred during the pervious 12 months, up to the end of the month preceding the date of interview (the reference period). In some parts of the survey respondents may also be asked about events that they have 'ever' experienced, for example some of the self-complete questions on drugs and sexual victimisation. This can lead to problems with respondent recall: some people may be less able to recall events accurately over such a long time-period, while others may mention things that happened outside the relevant time frame.

There are two main types of recall error, telescoping and omission, both of which are likely to occur in crime surveys (Lynn and Elliot 2000). Telescoping occurs when respondents report events, often highly salient or memorable, that occurred outside the reference period as if they occurred *inside* the reference period. To address this problem, respondents are asked to recall the month or, if necessary, the quarter in which the incident occurred. Omission occurs when respondents forget events, perhaps less salient or memorable, that occurred inside the reference period. This is harder to check with respondents.

It is highly likely that the point estimates produced by the SCJS are biased to some extent by issues of recall, although as long as recall bias is constant over time comparisons across different time periods are unlikely to be affected.

1.3.4 Unwillingness to report

Respondents may be unwilling to disclose some victimisation experiences in a face-to-face interview setting, particularly in relation to crimes such as domestic violence, rape and sexual assault. The self-completion modules are used to collect sensitive information and attempt to alleviate this problem. Work on the BCS has found that prevalence rates for domestic violence are around five times higher for self-completion modules than those obtained from the victim form (Nicholas et al 2007).

1.3.5 Definitions of crime

Incidents which are legally defined as offences may not be seen as such by respondents, who may then not report them to the survey. This definitional problem is particularly relevant for minor incidents and some forms of violence. In an attempt to overcome this problem, the SCJS victim form screener questions ask respondents whether they have experienced certain types of events. These questions do not refer to 'crimes', use legal terminology or refer to specific offences, for example 'Since [Insert date], has anyone GOT INTO your home without permission and STOLEN or TRIED TO STEAL anything?' If the respondent indicates they have experienced an event of this type, information is collected by means of the victim form and the coding process will determine whether the police would have defined this as a crime or not.

Overall, methodological studies suggest that victimisation surveys tend to undercount some trivial offences (e.g. vandalism and minor assaults) and sensitive offences (e.g. sexual offences and domestic violence), although less so than police statistics. More serious incidents that are not so sensitive, such as housebreaking, may be over counted, as more salient events tend to be telescoped into the reference period (see 'Recall' above).

1.3.6 Other definitional issues

Coverage

The SCJS is designed to be "a survey of public experiences and perceptions of crime in Scotland" (Page et al 2010b: 8). However, the survey is restricted to collecting data on a certain number of crime types. Thus there are restrictions on what 'crime' and 'victimisation' are according to the SCJS (and on their estimated prevalence in Scotland) – when talking about overall estimates of crime, for example, it is better to say 'all SCJS crimes' than 'all crimes', since a range of crimes are not included in the SCJS estimates (see 'Comparability' below).

Another key restriction is that the population covered by the survey is individuals aged 16 or over living in private residential accommodation in Scotland (all but the smallest islands are included). This means that people living in institutional settings, such as prisons, hospitals, army barracks, student halls of residence, nurse's accommodation and air bases, are excluded from the sample.

Those aged 15 or under are also currently excluded from the sample (although there were surveys of younger respondents built into the 1996 and 2000 Scottish Crime Surveys). Equally, businesses and non-commercial organisations are excluded, and the SCJS by definition only covers crimes experienced by individuals, not those experienced by organisations. Finally, victimless crimes (for example speeding) and crimes where the victim cannot be interviewed (such as homicide) are also excluded.

The SCJS covers only incidents of victimisation that occurred in Scotland. Incidents that affected people living in Scotland but which occurred in England and Wales are in reality missed altogether because Scottish households are not sampled by the BCS. Incidents that happened outside Great Britain are likewise excluded.

In addition to the excluded types of victim and crimes outlined above, the survey has a precise definition of the crime-types included in the estimates of 'all SCJS crimes' produced by the Scottish Government. Sexual offences are excluded because of the likely unreliability of reporting of these types of crime; and threats are excluded because it can be difficult to ascertain whether a crime actually occurred. Despite their exclusion from overall estimates, data on sexual offences and threats are collected in the victim forms, and are available for analysis by data users.

Comparability with police statistics

Crime surveys in the UK were established specifically to act as an alternative measure of crime to the police recorded crime statistics, which were recognised as providing only a partial measure of crime. While it has never been claimed

that crime surveys provide a completely accurate measure of crime, they do provide a better measure than police statistics for some crime types because they include crimes that are not reported to the police and crimes that are reported to but not recorded by the police. Crime survey statistics can only be compared against police recorded crimes for some categories of crime (known as the 'comparable subset'), however. In the SCJS, there are six offence groups that can be compared to police data. These are: vandalism, housebreaking, theft of a motor vehicle, bicycle theft, assault and robbery.

The comparisons are based on the fact that these crime types can be reasonably accurately coded in the same way as the police would do (as there is less likely to be ambiguity around these crime types as to what actually happened). However, there are also limitations to the comparisons. For example, the SCJS excludes respondents under the age of 16 and those not living in private households, whereas these victims would be measured in police statistics. Police statistics have to be carefully adjusted to match to SCJS data, so that victims under the age of 16 and crimes against commercial premises are excluded. However, this is not a completely problem-free process. Excluding victims who do not live in private residences is virtually impossible to do. Further, it has not always been possible to get adjusted data for the whole of Scotland (in previous surveys grossed up data based on police statistics for the Strathclyde police force area have been used). According to the Technical Report for the 2008/09 survey, "various adjustments were made to the recorded crime categories to maximise comparability with the SCJS" (Page et al, 2009:96); however, the data were not adjusted to remove crimes against victims under the age of 16 or crime against commercial businesses. This is consistent with established practice for the British Crime Survey.

Since 2004, Scottish police have conformed to a new set of Police Crime Recording Standards which were introduced in an effort to provide a more victim-centred means of recording crime. According to this set of Standards, there should be far less discrepancy between what crimes are reported to the police and those that are recorded by the police. In the crime survey reports for Scotland covering the period from 1993 to 2004, data was provided that demonstrated the gap between crimes recorded by the police, crimes reported to the police by victims and crimes that were experienced but never reported to the police. These data highlighted a large discrepancy between these numbers. However, since the introduction of the new police Standards in 2004, crime survey reports have not reproduced these data; therefore, it is not possible to determine whether the Standards have closed the gap between what members of the public report to the police and what the police actually record. It should also be noted that the introduction of the new rolling reference period used in the 2008/09 SCJS means that the period over which individuals were victimised spans almost two years and does not exactly map to the police recorded crime statistics, which cover a fixed financial year reference period.

1.3.7 Geographic breakdown

A number of different geographic groupings based on administrative boundaries within Scotland are provided in the SCJS data files. This includes Police Force Area (PFA), Community Justice Authority Area (CJAA), national Criminal Justice Board Area (NCJBA), Local Authority Area (LAA) and Health Board Area (HBA). From a user's perspective, this means that aggregate data from the SCJS at any of these levels could be merged onto other datasets for analysis. For example, percentage of the population who are victimised could be calculated for health board areas, and merged on to health board level data. Equally, data at each of these levels could be merged onto the SCJS. For example, mortality rates at local authority level could be merged onto the SCJS data. It should be noted, however, that the sampling strategy used for the survey means that the data are only really designed to be representative at PFA and all Scotland levels.

The SCJS data files released to the ESDS Archive do not hold small area units of geography (such as postcode or datazone) that would allow for detailed geographical analysis. However, for users wishing to conduct more complex analysis of the data, including multi-level modelling, it is possible to apply to the Scottish Government Justice Outcomes Analytical Unit for a data access agreement for a specially restricted dataset that includes a geographical marker. For reasons of confidentiality, it is only possible to get information on the Intermediate Geography area in which the respondent was living.

Section 2: Using the data

2.1 Introduction

This section of the User Guide is intended to provide practical help and advice on gaining access to and analysing data from the 2008/09 Scottish Crime and Justice Survey. Most of the information provided here will also be relevant to the 2009/10 survey and subsequent datasets, although users should note that there are likely to be changes to the survey questionnaire, datasets and design over time.

In this section, we identify how to gain access to the data, explain the format and content of the SCJS data files and how to select the appropriate one for your analysis, provide hints and tips on data management issues encountered when using these data, explain what weights to use when analysing the SCJS data and why different weights are necessary and illustrate how to merge variables from one dataset to another. In addition, this guide provides specific examples of analysis that can be conducted using the data to replicate findings from the published report (MacLeod et al 2009). Examples of syntax for SPSS are provided, and it is intended that users adapt this syntax as necessary to conduct their own analysis.

Please note that this User Guide is intended to be read in conjunction with, and not instead of, the SCJS Technical Report (Page et al 2009) which provides important additional information about the construction of the data files.

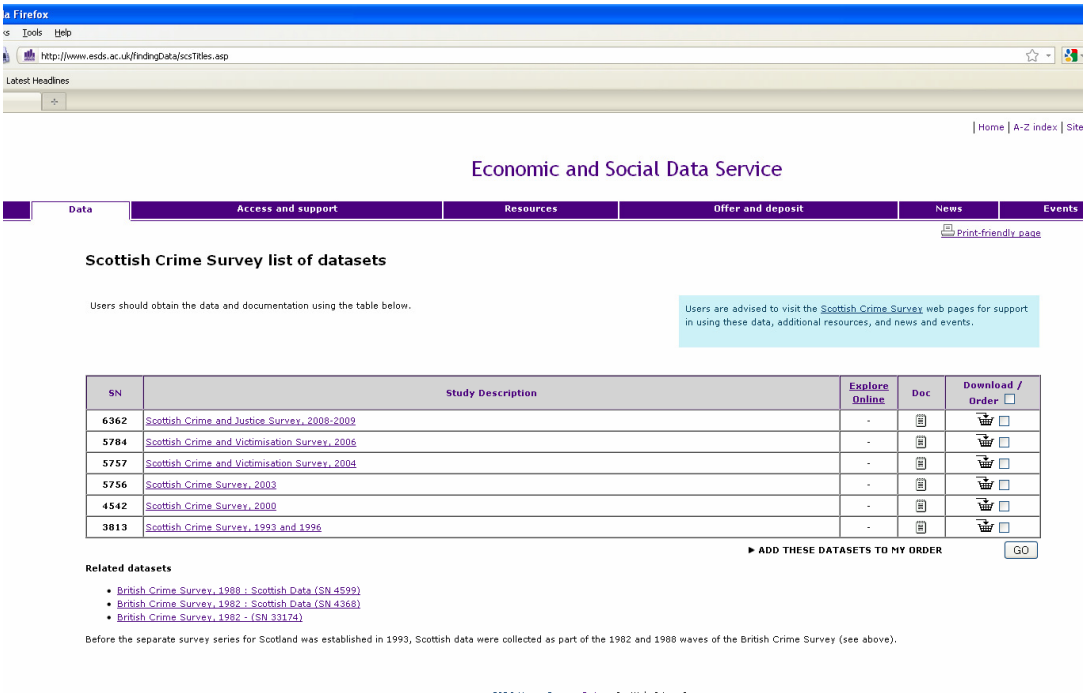
2.2 Access

SCJS data files can be downloaded from the SCJS page on the Economic and Social Data Service (ESDS) website: http://www.esds.ac.uk/findingData/scsTitles.asp. Data can be downloaded free of charge for academic users, but note that users are required to register with ESDS and give a brief explanation of why they want to use the files before data can be downloaded. Figure 2.1 provides a screenshot showing the ESDS download page. Data are currently available for the 2008/09 survey and all of the previous crime surveys in Scotland.

In addition to the raw data files, a large number of computer tabulations highlighting headline figures from the 2008/09 SCJS and cross analysed by a range of demographic, geographic, attitudinal and experiential variables are available from the Scottish Government website by using the link below. These tables may be useful to researchers both in terms of already providing the data they are interested in and also checking analysis undertaken in SPSS. This site also provides a series of data tables for the 2009/10 SCJS.

http://www.scotland.gov.uk/Topics/Statistics/Browse/Crime-Justice/Datasets/SCJS.

Figure 2.1: Screenshot of the ESDS website download page for SCJS data



2.3 File format

Both SPSS and Stata data files can be downloaded from the ESDS website. The syntax provided in this user guide is for SPSS (or PASW) users; however, all of the examples could be run equally using the appropriate code in any other statistical package and should provide the same results.

Three data files from the 2008/09 SCJS are available from the ESDS Archive:

- the respondent file or RF (filename: scjs_s2_rf_091214);
- the victim form file or VFF (filename: scjs_s2_vff_091210);
- the self-completion file or SCF (filename: scjs_sc_091209).

Section 1.2 of this User Guide, and chapter 3 of the Technical Report (Page et al 2009: 22-43), provide a detailed description of the SCJS questionnaire design, content and coverage. It is highly recommended that before analysing the data, users familiarise themselves with the survey questionnaire (this is also available to download from the ESDS website). The SCJS is a highly structured survey which contains multiple routing, which can be observed from the questionnaire but is not always obvious from the data files.

2.4 Levels of analysis

Most social surveys allow for analysis at one level (usually to make statements about respondents and extrapolate to the wider population). However, the UK crime surveys allow for statements to be made at three different levels: the level of the individual respondent, the level of the respondent's household and the level of specific incidents of crime. Each of these levels is discussed in more detail below. It is important for data users to note these differences in unit of analysis, as it has implications for how the data should be weighted (see section 2.6) and how the results of the analysis should be presented.

2.4.1 Individual level analysis

Most users of the SCJS data will probably want to make general statements about the characteristics, attitudes or experiences of the respondents, so as to extrapolate to the wider Scottish population. This involves conducting individual level analysis, either using the respondent file (RF) or the self-completion file (SCF) to produce point estimates.

2.4.2 Household level analysis

The design of the survey involves sampling households as the first stage, and then sampling people within households as the second stage. Therefore, it is possible to conduct some analysis at household level (this can also be done using the RF or the SCF). This approach might be taken in two main contexts. The first would be to analyse variables that relate specifically to the characteristics of the household (e.g. household tenure or property type). The second would be to conduct analysis of household crimes (i.e. those that affect all the members of the household), rather than personal crimes (i.e. those which affect individuals only). Analysis at the household level requires a different set of weights to individual level analysis, and this is explained in section 2.6.

2.4.3 Victim incident level analysis

The main reason for conducting the SCJS is to collect information from victims about their experiences of crime. Because detailed information is collected about all incidents of victimisation (up to a maximum of 5 victim forms), the data has a hierarchical structure. The victim form data file (VFF) allows analysis at the level of the crime incident because each case relates to one single crime or series of crimes (this is discussed further in section 2.5). As most users are interested in examining the characteristics of crime victims, information from the VFF is aggregated up to individual level and certain key variables are saved into both the RF and the SCF. However, those who wish to explore the characteristics of individual crimes can do so using the VFF. For example, you may wish to explore the proportion of incidents of specific crime types which involved the use of a weapon or that were reported to the police. Again, a different set of weights are required for this type of analysis, as discussed in section 2.6.

2.5 Choosing a dataset

Choosing the appropriate dataset for use will depend on the nature of the users' research questions and where, within the survey, the questions of interest were located. The contents of the three data files are discussed in a little more detail below; however, it is important to recognise that users can combine datasets together in a range of different ways to suit users' purposes. Merging the datasets is discussed in section 3.4, below.

2.5.1 The respondent file (RF)

The RF file is produced at the level of the individual respondent and contains data for all respondents, regardless of whether they are victims or non-victims. This is probably the data file most commonly used by data users because it contains primary data from all the questions contained in the main questionnaire, full sample modules, quarter sample modules and the demographic information about the respondents (see Figure 1.1 on page 10 of this User Guide). In addition, the RF contains useful variables drawn from administrative data sources, including a range of geo-demographic indicators (such as Police Force Area, Local Authority Area and Community Justice Authority Area) and Scottish Index of Multiple Deprivation indicators. Very helpfully, the survey company also provide a number of derived variables which have been drawn from the victim form file (VFF). This allows users to conduct analysis of victimisation at the individual level (i.e. to explore the characteristics and attitudes of victims of crime), without first having to aggregate the data contained in the VFF.

The derived variables contained in the RF include five 'VICFORM' and three 'VICFLAG' variables which provide some technical information relating to the VFF. VICFORM variables denote the offence code assigned to each Victim Form (each respondent can have up to five forms). The VICFLAG variables indicate the individual's victim status: whether they reported being a victim at all (VICFLAG1); whether they were a victim of a 'valid' offence (i.e. an offence that falls within the scope of the survey, that occurred within the reference period and happened within Scotland) (VICFLAG2); and whether they were a victim of a valid SCJS offence (i.e. excluding sexual offences and threats) (VICFLAG3). These variables can be used if users are simply interested in whether the individual was a victim or not, and wish to combine that information with other data collected within the main part of the survey questionnaire.

A further set of derived variables are provided for the individual crime types or groups. This means that if you wish to conduct more specific analysis by crime type, these variables allow you to do so without having to create them yourself from the VFF. Table 2.1, below, provides a list of the secondary variables contained in the 2008/09 RF, including prevalence variables (i.e. whether or not the respondent was a victim of that crime type), incidence variables (i.e. how

many incidents of that type of crime the respondent experienced)[5] and repeated prevalence variables (i.e. whether or not the respondent was a victim of that crime type on more than one occasion). A description of the variables is provided; however, the nature of the crime types and groups is relatively easy to understand from the variable names (e.g. PREVMOTOVVAND is prevalence of motor vehicle vandalism); however, detailed labels are also provided in the RF dataset so that users can determine which variables they wish to use. Tables A1, A2 and A3 in the Appendix to this guide also provide useful information about the specific crime categories that are amalgamated to create these crime variables.

Repeat prevalence variables are not provided for all crime types; however, it is a simple procedure to create other variables. For example, if you wish to create a repeat prevalence variable for all assaults (and not just serious assaults, as contained in the dataset), you could do this by recoding the variable for incidence of assault (INCASSAULT) using example syntax 1. This syntax recodes INCASSAULT such that all respondents who have experienced two or more incidents of assault are given a value of 1, while those who are non-victims or have experienced only one incident of assault are given a value of 2. The 'var label' and 'val label' commands provide labels for both the new variable and its two values (1 and 2).

Example syntax 1: Creating repeat victimisation variables

```
recode INCASSAULT (2 thru hi=1) (else=2) into REPASSAULT.
var label REPASSAULT 'REPASSAULT: Repeat victim of assault'.
val label REPASSAULT 1 'Yes' 2 'No'.
```

Finally, the RF contains six derived variables, also drawn from the VFF, which identify whether the respondent was a victim of *any crime* where the offender was under the influence of either alcohol or drugs (ALCOFF and DROFF), and also whether the respondent was a victim of a *violent crime* in which they (ALCVIOVIC and DRVIOVIC) or the offender (ALCVIOOFF and DRVIOOFF) was under the influence of drink or drugs Respectively.

Weights are also contained in the RF, and these are discussed in section 2.6.

[5] This incidence variable is affected by a cap of 5 that is attached to series incidents during the weighting process – see section on 'the victim form file'.

Table 2.1: Derived crime variables contained in the SCJS 2008/09 respondent file

Variable description		Prevalence variable	Incidence variable	Repeat prevalence variable
Household crimes	Motor vehicle vandalism	Prevmotovvand	Incmotovvand	Repmotovvand
	Property vandalism	Prevpropvand	Incpropvand	Reppropvand
	Theft from a motor vehicle	Prevtheftfrommv	Inctheftfrommv	
	Attempted theft of or from a motor vehicle	Prevatttheftmv	Incatttheftmv	
	Theft of a motor vehicle	Prevtheftofmv	Inctheftofmv	
	Bicycle theft	Prevbicycletheft	Incbicycletheft	
	Housebreaking	Prevhousebreak	Inchousebreak	Rephousebreak
	Other household theft	Prevotherhousetheft	Incotherhousetheft	
Personal crimes	All assault (serious and minor assault)	Prevassault	Incassault	
	Serious assault	Prevserassault	Incserassault	Repserassault
	Sexual assault	Prevsexoff	Incsexoff	
	Theft from the person	Prevtheftfperson	Inctheftfperson	
	Robbery	Prevrob	Incrob	Reprob
	Other personal theft	Prevothertheft	Incothertheft	
	Threats	Prevthreat	Incthreat	
Combined crime categories	Vandalism	Prevvand	Incvand	Repallvand
	All household crime	Prevhouse	Inchouse	
	All personal crime	Prevperson	Incperson	
	All property crime	Prevproperty	Incproperty	Repproperty
	Comparable subset of crimes	Prevcomparcrime	Inccomparcrime	
	Acquisitive crime	Prevacquis	Incacquis	
	Violent crime (assault and robbery)	Prevviolent	Incviolent	Repviolent
	All SCJS survey crime	Prevsurveycrime	Incsurveycrime	
	All motor vehicle thefts	Prevallmvtheft	Incallmvtheft	Repallmvtheft
	Other household theft including bicycle theft	Prevotherhousetheftcycle	Incotherhousetheftcycle	Repotherhousetheftcycle
	All personal theft	Prevperstheft	Incperstheft	Reppersttheft

2.5.2 The victim form file (VFF)

The VFF contains the data collected in the victim forms (up to 5 per respondent) and is produced at the level of the crime incident. Because each respondent may be represented more than once in this data file (i.e. completed more than one victim form), the number of cases far exceeds the number of actual victims. This means that the unit of analysis for this dataset is crime incidents, and not individual victims. Data users who wish to use the victim form file may be interested in looking at the characteristics of crimes, rather than the characteristics of their individual victims.

The VFF contains information about every incident of victimisation reported by a respondent to the survey; however, not all of these cases are used to calculate the official victimisation rates published in the First Findings report (see MacLeod et al 2009). This is because only those incidents that have occurred within the relevant reference period (12 months prior to the month of interview) or inside Scotland are considered 'valid' SCJS crimes, and suitable for comparison with the police recorded crime statistics. In addition, any crimes that are defined as 'outside the scope of the survey' are also counted as 'invalid' (for further information on invalid and out of scope crimes see chapter 7 of the Technical Report, Page et al, 2009: 66-68).

A weight is created which sets the value of all invalid or out of scope crime incidents to zero for the purposes of analysis in the VFF data file. It is important to note that data users who wish to analyse all incidents of a particular type of crime within the VFF data file (e.g. assault), including those that were not classified as valid SCJS crimes, would need to recreate the prevalence and incidence variables shown in Table 2.1. Further information on how to do this is provided in section 2.6, below. Please note, however, that as these cases are not considered essential to the survey, very little information is collected about them as the victim form is terminated prior to completion.

The VFF starts with a series of offence code variables which identify how the incident has been coded by different sources (this is a checking mechanism to ensure that crimes are coded accurately, see section 1.2.4). The variable of principal interest to data users is OFFENCE which provides a final offence code for each victim form. Note that during the coding process only one offence code is assigned; therefore, if a crime involved two components (e.g. a housebreaking and an assault), the most serious of these two components (in this case the assault) would be given priority.

The variable OFFENCE provides a very specific code for the type of crime experienced by the respondent, which are carefully allocated so as to reflect as far as possible the way in which the incident would have been coded by the police. However, most users will probably wish to use a variable that combines these offence codes into broader crime types (e.g. all assault or all motor vehicle thefts). Therefore, the OFFENCE variable can be recoded to combine crime codes together. Users can, of course, do this in any way that they wish. However,

syntax example 2 provides the SPSS command to create a new variable, OFFTYPE, which mirror the SCJS offence types contained in the RF.

Example syntax 2: Combining individual offence codes to form broad SCJS crime types

```
compute OFFTYPE =98.
if (offence=49) OFFTYPE =1.
if (offence=48 or offence=50 or offence=51) OFFTYPE =2.
if (offence=35 or offence=37) OFFTYPE =3.
if (offence=44 or offence=45) OFFTYPE =4.
if (offence=34 or offence=36) OFFTYPE =5.
if (offence=38) OFFTYPE =6.
if (offence=25 or offence=26 or offence=27) OFFTYPE =7.
if (offence=24 or offence=29 or offence=30 or offence=31 or offence=32
or offence=39) OFFTYPE =8.
if (offence=3 or offence=7) OFFTYPE =9.
if (offence=2 or offence=4 or offence=5) OFFTYPE =10.
if (offence=8 or offence=9 or offence=10 or offence=11 or offence=12 or
offence =13 or offence=14 or offence=15) OFFTYPE =11.
if (offence=19 or offence=20 or offence=21) OFFTYPE =12.
if (offence=17 or offence=18) OFFTYPE =13.
if (offence=41 or offence=46) OFFTYPE =14.
if (offence=55 or offence=56 or offence=57 or offence=58) OFFTYPE =15.
if (offence=1 or offence=6 or offence=16 or offence=22 or offence=23
  or offence=28 or offence=33 or offence=40 or offence=42 or offence=43
or offence=47  or offence=52 or offence=53 or offence=54 or offence=59
or offence=60 or offence=61   or offence=62 or offence=63) OFFTYPE =16.

var lab OFFTYPE 'Type of offence recorded for each incident of
victimisation'.
val lab OFFTYPE 1 'mv vandalism' 2 'property vandalism' 3 'theft from
mv' 4 'att theft of/from mv'  5 'theft of motor vehicle' 6 'bicycle theft' 7
'housebreaking' 8 'other household theft'   9 'minor assault' 10 'serious
assault' 11 'sexual assault' 12 'theft from person' 13 'robbery'   14 'other
personal theft' 15 'threats' 16 'other crimes and incidents non-classifiable
or outside the scope of the survey'.
```

Another key variable of interest in the VFF is PINCI. This variable shows whether the respondent reported one 'single' incident (e.g. a one-off assault) or a 'series' of incidents (e.g. several related assaults). In common with other crime surveys, a series of incidents is defined as "the **same thing**, done under the **same circumstances** and **probably by the same people**". This might include a series of violent domestic assaults or a series of vandalism incidents committed by local youths. A significant minority (22%) of the victim forms collected by the survey were for series incidents; therefore, these are well worth analysing. However, it should be noted that during the weighting process, victim forms detailing more than five incidents of any one crime type are given a 'cap' of five. This is done to prevent over-inflation of the crime estimates and broadening of the confidence

intervals when the figures are grossed up to population level. However, it restricts the opportunities for analysis for those interested in studying repeated victimisation. Therefore, data users may wish to recreate the 'incidence' variables contained in the RF data file but without this cap. Further details on how this can be done are given in section 2.6.

The VFF provides detailed information on the nature of each offence (or, in the case of series incidents, on the most recent offence). This includes: when and where it occurred; possible motives of the offender; specific information about how each type of offence was committed (i.e. modus operandi); details of what or who was affected by the crime; (where known) information about the characteristics of the offender; the use of weapons; emotional response to the crime; whether it was reported to the police (and reasons why if it was not); support received by a range of agencies; how far the crime progressed through the criminal justice system; levels of satisfaction with the police and other agencies; and opinions about what did or should have happened to the offender.

As with the RF, some useful geo-demographic variables are appended to the data file (including Police Force Area, Local Authority Area and Community Justice Authority Area) and two Scottish Index of Multiple Deprivation indicators.

Three weights are contained in the VFF. These are discussed in section 2.6 below; however, it is worth noting that one of these weights, WGTGINC_SCJS, is unique to the 'crime incident level' data. This is the weight that sets all non-valid and out of scope cases to zero and caps the series incidents forms to a maximum of five.

2.5.3 The self-completion file (SCF)

Like the RF data file, the SCF data file is produced at the level of the respondent. This data file contains the data collected from the self-completion questionnaire (including questions on illicit drug use, stalking and harassment, partner abuse and sexual victimisation). In addition, it includes the same key demographic variables, geo-demographic indicators, SIMD variables and the derived crime variables that are found in the RF data file. The SCF can easily be linked to the RF data file (if, for example, data users wished to merge different variables from each of these components of the survey together). The two datasets can be matched using the variable SERIAL (see section 2.6 for details of how to merge data files). However, it is important to note that the two files do not contain the same number of cases. This is because not all of the respondents to the main component of the survey also took part in the self-completion component. For this reason, different weights are required to conduct any analysis that uses data from the self-completion questionnaire. Please also note that in the illicit drugs section of the SCF data file, the responses from respondents who reported taking a bogus drug 'semeron' are not usually included in analysis.

2.6 Selecting the appropriate weight

All large-scale sample-based surveys use weights to adjust for bias in the design of the survey, the sampling process and the response rate. It is important for users to consider these issues when analysing the data, as it will be necessary to use weights in order to produce the most accurate estimates. The SCJS (and other crime surveys) tend to be more complex than other national surveys in terms of the choice of weights supplied in the datasets, mainly because of the multifaceted nature of the survey. A detailed discussion of the weights can be found in chapter 8 of the Technical Report (Page et al 2009: 77-86). Briefly, the calculation of weights for the SCJS takes account of three key things:

- Rates of response vary for different sub-groups in the sample. Response varies by age, gender and household type, so non-response corrections or 'calibration weighting' is required.
- Probabilities of selection are unequal due to the sample design. Corrections to the sample need to be made to account for: probability of selection for respondents in different sized households and the requirement to sample at least 1,000 adults in each Police Force Area. These are called 'design weights'.
- An expansion factor is used to create grossing weights so that data can be expressed at population level, for example to provide national estimates of victimisation. Scaled versions of the household and individual weights are also provided, which do not include the expansion factor.

The three 2008/09 SCJS data files contain a total of nine weights. Once you have selected the data file you wish to use, you will be restricted to the weights contained in that dataset (however, these are somewhat overlapping, so it is useful to have a broad overview of all the weights). Table 2.2, below, shows the variable names of the nine weights, indicates in which datasets they can be found, and gives a brief description of each weight provided in the 2008/09 SCJS datasets.

Table 2.2: Weighting variables provided in the 2008/09 SCJS data files

Variable name	Data file	Description of weight
WGTGHHD	RF & VFF	Grossed household weight
WGTGINDIV	RF & VFF	Grossed individual weight
WGTGINC_SCJS	VFF	Grossed incident weight for SCJS crimes
WGTGHHD_SC	SCF	Grossed self-completion household weight
WGTGINDIV_SC	SCF	Grossed self-completion individual weight
WGTGHHD_SCALE	RF & VFF	Scaled household weight
WGTGINDIV_SCALE	RF & VFF	Scaled individual weight
WGTGHHD_SC_SCALE	SCF	Scaled self-completion household weight
WGTGINDIV_SC_SCALE	SCF	Scaled self-completion individual weight

As discussed in section 2.4, different weights are required depending on the unit of analysis. There are three units of analysis, and the weights that should be applied are as follows:

- Weights containing the suffix 'HHD' are for 'household level' analysis. These should be applied when users wish to conduct analysis of crimes or characteristics relating to the whole household.
- Weights containing the suffix 'INDIV' are for 'individual level' analysis. These should be applied for the analysis of personal crimes or where users wish to explore the characteristics or opinions of the individual respondent.
- The weight containing the suffix 'INC_SCJS' is for 'crime incident level' analysis. This should be used for the analysis of crime incidents rather than victims of crime.

Household and individual level weights are available in all three of the SCJS datasets; however, slightly different weights are provided for the SCF because the sample base is different to that of the respondent file (as not all respondents participated in the self-completion element of the survey). The crime incident weight is only contained in the VFF.

Users also have the choice of using 'grossed' or 'scaled' weights. The grossed weights involve inflating the sample size to the number of households or individuals in Scotland. These should only be used if you wish to calculate national estimates for either household or personal crimes. For more general analysis of the sample data, data users should always use the scaled weights, which have the suffix _SCALE. Scaled weights are adjusted so that the weighted sample 'n' is the same as the unweighted 'n'. In practice, using the grossed or scaled weight should make no difference in terms of actual estimates or percentages in your results; however, using the grossed weight will greatly increase the probability of a type I error in any statistical tests that you run i.e. results will appear to be statistically significant when they are not.

Users should note that there is no scaled incidence weight contained in the VFF. The incidence weight is calculated by multiplying either the grossed household weight or the grossed individual weight by a further inflating factor which takes account of the number of incidents of crime represented by each victim form completed by a respondent (to a maximum of five incidents per form). The grossed weight used to calculate the incidence weight is dependent on whether the victim form represents a household or a personal crime. The incident inflating factor has values ranging from 0 to 5, where a value of 0 represents a victim form that is not a valid SCJS crime (i.e. it is a crime that does not fit within the standard SCJS crime categories, or occurred outwith Scotland or outside the relevant 12 month reference period). This factor is the same regardless of whether the victim form represents a household or a personal crime.

2.6.1 Applying the weights

Weights can be applied using the menus in SPSS, by going to the 'Data' option and clicking on 'weight cases' at the bottom of the list. Simply enter the variable name for the weight you wish to apply into the 'weight cases by' dialogue box. Alternatively, a simple piece of syntax can be used to turn on weights, as shown in example 3 below. In addition, it is important to remember to turn off weights at other times (for example, when you wish to report the raw 'n' for your analysis), so syntax for this is also provided.

Example syntax 3: Turning on and off weights

```
weight by WGTGHHD.

weight off.
```

Further examples of using weights to conduct specific pieces of analysis are provided in section 3 of this guide.

2.6.2 Calculating different weights

Additional weights can also be calculated to conduct comparative analysis with previous surveys. For example, if you wish to examine incidence of crimes as a rate per 10,000 households or individuals (rather than national estimates) this can be done by dividing the gross weights by the total population (of either households or adults aged 16 or over in Scotland) divided by 10,000. Population estimates provided by the General Registrar Office for Scotland shows that in 2008, there were 2,331,250 private households in Scotland and 4,254,966 adults aged 16 or over living in Scotland (see Page et al 2009: 69). Example syntax 4 shows how to create new weights that can be used to calculate victimisation rates per 10,000 households (WGTHHDRATE) and individuals (WGTINDIVRATE) for the 2008/09 SCJS.

Example syntax 4: Computing new weights to calculate victimisation rates

```
compute WGTHHDRATE=WGTGHHD /(2331250/10000).
compute WGTINDIVRATE=WGTGINDIV/(4254966/10000).
```

It is also possible to create weights to calculate victimisation rates for motor vehicle crime relating only to households in Scotland that owned a vehicle or for bicycle crime relating only to households in Scotland that owned a bicycle. It is difficult to find objective data sources that can provide information on the number of households that own motor vehicles or bicycles in Scotland, so the data provided by the SCJS respondents is relied on to provide this information. Using the grossed household weight, create an estimate for the number of households that owned a motor vehicle or a bicycle (using variables MOTORCYC, CAR and OWNBIK2 in the respondent data file) and then use the same syntax as example 4 to create the appropriate weight.

To illustrate this example, syntax 5 shows how to create a weight to calculate crime rates per 10,000 bicycle owning households. Start by applying the household weight and run a frequency to estimate the number of households in Scotland that owned a bicycle during the reference year (the answer is 868,680). Compute a new weight for calculating victimisation rates per 10,000 bicycle owning households (WGTBIKERATE), by dividing the grossed household weight by the number of bicycle owning households divided by 10,000. Apply this new weight and run a frequency analysis of the bicycle theft incidence variable (INCBICYCLETHEFT). It is important at this stage to only include bicycle owning households in the analysis (this can be done by filtering on the OWNBIK2 variable or using a temporary 'select if' command, as has been used in this example). Ask for a 'sum' on the frequency command and the results will show a value of 294. This means that 294 per 10,000 bicycle owning households had a bicycle stolen during 2008/09 (the same as the figure published in the SJCS 2008/09 report, MacLeod et al 2009: 177).

Example syntax 5: Computing a new weight for bicycle owning households

```
weight by WGTGHHD.
fre OWNBIK2 .

compute WGTBIKERATE=WGTGHHD /(868680/10000).

weight by WGTBIKERATE.
temporary.
select if OWNBIK2 eq 1.
fre INCBICYCLETHEFT/for not/sta sum.
```

2.7 Checking and defining variables

Before starting to conduct your analysis, users should spend a little time checking the variables contained in the relevant data file. This can be done in a number of ways. Firstly, it is useful simply to observe the structure of the data using the 'variable view' window in SPSS. The data in each of the three files are ordered in the same way as the questionnaires, which makes it useful to have the questionnaires to hand when analysing the data.

The second way to observe the structure of the data (i.e. to see how variables and values are labelled) is to run a data dictionary. This can be done by going to 'File' in the menus, clicking on the option 'display data file information' and choosing 'working file'. Two tables are automatically pasted into the output file: the first table is headed 'variable information' and contains basic information about the variable name, label, measurement type and what values are set to missing; the second table is headed 'variable values' and this provides information about the active values for each variable in the dataset (i.e. it does not provide information on variables that are set as missing).

A third way to explore the data is to run some simple frequencies of the variables you are interested in. Frequency analysis allows you to observe the distribution

of the values in each variable. You can use this to check whether there are values that you may wish to re-define before analysis begins. For example, you may wish to exclude some values that you are not interested in by setting them to missing (e.g. 'don't know' or 'refused') or recode some variables to combine categories. Recoding should be done where you wish to retain the old variable (as a point of reference), but also create a new variable.

2.7.1 Setting missing values

Datasets usually have certain values that are pre-set to 'system missing', such as where an earlier filter in the questionnaire has been used and some respondents have been diverted away from a set of questions. However, it is also possible for users to set values as missing if they do not wish them to be included in the analysis. This is known as setting values to 'user missing', and it is important that users check variables to determine whether there are values that are not of interest to them. Examples in the SCJS 2008/09 data files may include values such as -2 for 'refused' or -1 for 'don't know' (although the decision to exclude 'don't know' responses should be carefully considered).

Values can be set to missing in two ways using SPSS. The first is to go to the 'variable view' window and look at the column headed 'missing'. Find the variable you are interested in and click on the box in the missing column. A small dialogue box with three dots will appear. Click on this dialogue box and a 'missing values' dialogue box will appear. This box will show any values that are already set to missing (e.g. 99). Type in any additional values that you want to be set to missing in the 'discrete missing values' boxes (up to 3 can be specified) or, alternatively, if you wish to make a range of values missing this can also be specified by setting the lowest and highest value. Then click OK. If there are whole blocks of variables that you wish to amend in the same way, you can simply copy and paste the values in the 'missing' column of the variable view window (this works in SPSS Version 17 and above).

The second way to set missing values is to use a simple piece of syntax, as shown in example 6 below. The example uses the variable QWORR_5 (How worried are you about your home being broken into?) which is taken from the RF data file, and shows how to set the values for 'refused' (-2) and 'don't know' (-1) to missing. Example 6 also shows you how to unset this command if you wish to include these values in the analysis again. Please note that this 'missing values' command only allows you to set up to 3 values as missing; if you wish to exclude four or more values from your analysis, it is necessary to recode your variable into a new one (see section 2.7.2).

Example syntax 6: Defining missing values and cancelling missing values

```
missing values QWORR_5 (-1, -2).

missing values QWORR_5 ( ).
```

2.7.2 Recoding variables

There are many reasons why it may be desirable to recode variables. If there are more than three values that you wish to set to missing values, and this cannot be done using the 'missing values' command shown above, recoding the variable may be the best option. Or you may simply wish to collapse or combine categories in your own particular way. Recoding can be done using the dialogue boxes by clicking on the 'Transform' button in the menu. You can do two types of recode operation: one which changes the original variable, and one which saves into a new variable (the second option is recommended as the safest option, as you do not lose any of the raw data). Recoding can also be done fairly simply using syntax.

Example 7, below, provides syntax showing how to recode the variable QWORR_5. This is an ordinal variable that measures how worried the respondent is about their home being broken into, and has four main values of interest: very worried, fairly worried, not very worried and not at all worried. This example both excludes the categories we are not interested in (i.e. don't know and refused) and collapses categories together (i.e. merges very worried with fairly worried, and not very worried with not at all worried). This time we create a variable (WORRY_HBREAK) which contains only two values, and the missing values are set to system missing (sysmis) rather than user missing. Unlike the 'missing values' command, the values that have been set to sysmis cannot be simply brought back into the analysis for the new variable – to do this would require another recode. When creating a new variable, it is always important to label the variable and all its values, so that you are clear what the data represents.

Example syntax 7: Recoding a variable to combine categories and exclude missing variables

```
recode QWORR_5  (1,2=1) (3,4=2) (else=sysmis) into WORRY_HBREAK.
variable labels  WORRY_HBREAK 'How worried are you that your home will be broken into?'.
value labels WORRY_HBREAK 1 'Very or fairly worried' 2 'Not very or not at all worried'.
execute.
```

Once you have created a recoded variable in this way, it is good practice to cross tabulate the old and new variables to check that the operation was performed correctly.

2.7.3 Derived variables

As noted in section 2.5 of this user guide, the SCJS data files already contain some derived variables that have been provided for the data user's convenience by the survey company. There are two main types of derived variable (Page et al, 2009: 88). The first are flag variables, which identify technical details such as the

quarter sample module the respondent was allocated to, the date of interview, whether the interview was full or only partial and whether the individual was or was not a victim of crime. On the victim form file, the flag variables indicate whether each form relates to a single incident or a series of related incidents and whether the incident was considered 'valid' for the SCJS. The second type of derived variables are classificatory variables which are derived from the data collected. This includes standard classifications of mainly demographic data, such as banded age groups, income groups, socio-economic group, ethnic group and tenure.

Some of these derived variables are created by combining two or more existing variables. For example, the variable TENURE has been created by combining the variables QDTENUR (tenure of home) with QDRENT (who is the property rented from). By combining these two variables, we have a more detailed variable which tells us whether the household tenure is owner-occupied, social rented, private rented or some other kind of tenure.

This type of derived variable can be created using the 'compute' command, which is also found in the 'Transform' menu in SPSS. Computing new variables allows you to create a bespoke variable by taking combinations of other variables and putting them together in your own way. This can be done using the dialogue boxes (which keeps you right, in terms of the logic required) or using syntax (which can be quicker, especially if you want to create several new variables, but take care to check that the logic of the command makes sense). Creating the first variable using the dialogue boxes and then pasting the syntax into a new window is a recommended way of replicating, and keeping a record of, your analysis.

The syntax used to create the TENURE variable is shown in example 8, below. You can run this syntax and then check it against the variable TENURE in the RF dataset – you should get the same answer.

Example syntax 8: Creating a derived variable for tenure from two existing variables

```
compute TENURE =0.
if (any (QDTENUR,1,2,3)) TENURE =1.
if (QDTENUR=4 and any (QDRENT,1,2)) TENURE =2.
if (QDTENUR=4 and any (QDRENT,3,4,5,6,7)) TENURE =3.
if (QDTENUR=5) TENURE =4.
if (QDTENUR=-1) or (QDTENUR=-2) or ((QDTENUR=4 and QDRENT=-2))
TENURE =-1.
variable labels TENURE 'Tenure type'.
value labels TENURE 1 'Owner occupied' 2 'Social rented' 3 'Private
rented' 4 'Other' -1 'refused or dont know'.
missing values TENURE (-1).
```

Section 3: Examples of commonly used analysis

The preceding section contained some examples of analysis, mainly as a means of illustrating issues around data management and file merging. The aim of this section is to provide further examples based on common types of analysis carried out using crime survey data based on the experiences of the authors of the report and of analysts within Scottish Government who use the data regularly to inform policy. This section does not provide an exhaustive list of examples; however, a range of analyses are illustrated to give users some suggestions of syntax that they may wish to use or adapt for their own purposes.

3.1 Examples using the respondent data file (RF)

3.1.1 What is the estimated number of household and personal crimes that occurred in Scotland during 2008/09?

Approach/analysis

One of the main purposes of the SCJS is to calculate an estimate of the number of crimes that occur in Scotland during a one year period. This can be done very simply by asking for a frequency command that includes a weighted sum. The variables used to calculate national estimates are victimisation variables that are derived from the 2008/09 victim form file but which are aggregated and saved into the respondent data file. Those that measure the incidence of each type of crime (and have the suffix INC in the dataset) are included in this analysis. (These are listed in Table 2.1 in the slides)

It is important to remember that household and personal crimes need to be calculated separately, as there are different weights that apply. For household crimes, the grossing weight WGTGHHD should be used as this provides an estimate of the number of crimes against all private households in Scotland; while for personal crimes, the grossing weight WGTGINDIV should be used as this provides an estimate of the number of crimes committed against all individuals age 16 or over in Scotland.

Syntax

The syntax shown below can be used to calculate the national estimates for selected household and personal crimes shown in Table A1.1 of the SCJS 2008/09 First Findings report (MacLeod et al, 2009: 174). Note that the weights used here are the grossing weights.

For the household crimes:

```
weight by WGTGHHD.
fre incvand incacquis inchousebreak inctheftofmv incbicycletheft
  /for not/sta sum.
```

For the personal crimes:

```
weight by WGTGINDIV.
fre incviolent incassault incrob
  /for not/sta sum.
```

Output

The 'sum' value shows the national estimate for the household crimes below:

		incvand INCVAND: Number of vandalism incidents	incacquis INCACQUIS: Number of acquisitive crime incidents	inchousebreak INCHOUSEBREAK: Number of housebreaking incidents	inctheftofmv INCTHEFTOFMV: Number of theft of motor vehicle incidents	incbicycletheft INCBICYCLETHEFT: Number of bicycle theft incidents
N	Valid	2331250	2331250	2331250	2331250	2331250
	Missing	0	0	0	0	0
Sum		350375.82	63656.82	25484.61	7423.61	30748.60

Similarly, the 'sum' value shows the national estimate for the personal crimes:

		incviolent INCVIOLENT: Number of violent crime incidents	incassault INCASSAULT: Number of assault with attempted assault incidents	incrob INCROB: Number of robbery incidents
N	Valid	4255000	4255000	4255000
	Missing	0	0	0
Sum		316589.62	296893.05	19696.57

The output above shows that there were an estimated 350,376 incidents of vandalism against private households in Scotland during 2008/09. There were also an estimated 316, 590 incidents of violent crime committed against Scottish adults aged 16 or over. The 'sum' data for the other household and personal crimes shown in this output would be interpreted in the same way.

3.1.2 How does experience of victimisation vary by age and sex?

Approach/analysis

It is possible to explore the extent to which victimisation varies by age and sex using both the incidence and prevalence variables for victimisation types in the respondent data file, and the variables QDGEN (for gender) and QDAGE (for age). There is also a derived age variable QDAGE2 which is banded into 10 broad age groups. The analysis below uses the variables PREVPERSON and INCPERSON, which relate to the prevalence and incidence of all personal crimes (i.e. violence and personal theft). There are a number of possible ways to conduct this analysis, so a range of syntax is provided.

Syntax

(1) Start by weighting the data using the scaled individual weight and ask for a frequency of the prevalence variable PREVPERSON and the incidence variable INCPERSON simply to observe the variables:

```
weight by WGTGINDIV_SCALE.
fre  PREVPERSON INCPERSON.
```

(2) Set the incidence value for those who had not experienced any personal crimes to missing, so as to restrict analysis of incidence to victims only (this is not essential, but it is more useful to look at incidence amongst victims only):

```
Missing values INCPERSON (0).
fre  INCPERSON.
```

(3) To explore differences in prevalence by gender, use the following syntax:

```
crosstabs
  /tables=PREVPERSON  by  QDGEN
  /format=avalue tables
  /statistics=chisq
  /cells=count column
  /count round cell.
```

(4) To explore differences in prevalence by age, you can use two main methods. The first is simply to do a cross-tabulation to find the difference in age bands between victims and non-victims, using the following syntax:

```
crosstabs
  /tables= QDAGE2 by PREVPERSON
  /format=avalue tables
  /statistics=chisq
  /cells=count row
  /count round cell.
```

(5) The second is to look at the mean and median age of those who were victims of personal crime compared to those who were not, using the following syntax:

```
means tables= QDAGE by PREVPERSON
 /cells mean count stddev median
 /statistics anova.
```

(6) The differences in age can be demonstrated graphically. The syntax shown below allows you to examine a boxplot for the relationship between personal victimisation and age.

```
examine variables=QDAGE by PREVPERSON
 /plot=boxplot
 /statistics=none
 /nototal.
```

(7) Finally, it is interesting to look at an error bar which shows the mean age for victims and non-victims. You can also break this down further to include a comparison of males and females, using the following syntax.

```
graph
 /errorbar(ci 95)= QDAGE by PREVPERSON by QDGEN.
```

Output

(1) The frequency table for PREVPERSON shows that 6% of respondents to the SCJS 2008/09 were victims of a personal crime.

prevperson PREVPERSON: Victim of personal crimes with attempted assault (prevalence)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00 Yes	962	6.0	6.0	6.0
	2.00 No	15038	94.0	94.0	100.0
	Total	16000	100.0	100.0	

The frequency table for INCPERSON shows that 94% of people had no incidents of personal crime (as expected having observed the frequency table for PREVPERSON). Of the remaining 6% of respondents, most of them had experienced only one incident of personal crime in the reference year; but a very small proportion of respondents had experienced more than one. As we are really only interested in those who experienced one or more incidents, we can set the 'zeros' to missing for this variable (see step 2).

incperson INCPERSON: Number of personal crime incidents

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	15038	94.0	94.0	94.0
	1.00	647	4.0	4.0	98.0
	2.00	161	1.0	1.0	99.0
	3.00	62	.4	.4	99.4
	4.00	37	.2	.2	99.7
	5.00	42	.3	.3	99.9
	6.00	6	.0	.0	100.0
	7.00	2	.0	.0	100.0
	8.00	4	.0	.0	100.0
	9.00	1	.0	.0	100.0
	10.00	1	.0	.0	100.0
	Total	16000	100.0	100.0	

(2) After setting the 'zeros' to missing, we can report from the table below that the majority (67%) of victims of personal crime experienced only one incident, although a small minority (around 6%) of individuals were victims of five or more incidents of personal crime. Since this is straightforward count data, you could also ask for a mean of incidence of personal crime using the 'descriptives' commands; however, it is bound to be very close to 1.

incperson INCPERSON: Number of personal crime incidents

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	647	4.0	67.3	67.3
	2.00	161	1.0	16.8	84.1
	3.00	62	.4	6.5	90.5
	4.00	37	.2	3.8	94.3
	5.00	42	.3	4.3	98.7
	6.00	6	.0	.6	99.2
	7.00	2	.0	.2	99.5
	8.00	4	.0	.4	99.8
	9.00	1	.0	.1	99.9
	10.00	1	.0	.1	100.0
	Total	962	6.0	100.0	
Missing	.00	15038	94.0		
Total		16000	100.0		

(3) A cross-tabulation of the variables PREVPERSON and QDGEN shows that 7.4% of males and 4.8% of females were victims of a personal crime. A chi-square test (not shown) revealed that the gender difference in prevalence was statistically significant (p<.001).

prevperson PREVPERSON: Victim of personal crimes with attempted assault (prevalence) * qdgen QDGEN: Respondent's Gender Crosstabulation

			QDGEN: Respondent's Gender		Total
			1 Male	2 Female	
PREVPERSON: Victim of personal crimes with attempted assault (prevalence)	1.00 Yes	Count	565	397	962
		% within QDGEN	7.4%	4.8%	6.0%
	2.00 No	Count	7078	7960	15038
		% within QDGEN	92.6%	95.2%	94.0%
Total		Count	7643	8357	16000
		% within QDGEN	100.0%	100.0%	100.0%

(4) The output for the cross-tabulation between age bands and prevalence shows that younger people (especially those between 16 and 24) are most likely to have been victims of personal crime; whereas those in the older age bands, especially over 65, are least likely to have experienced this type of crime.

qdage2 QDAGE2: Respondent's age in bands * prevperson PREVPERSON: Victim of personal crimes with attempted assault (prevalence) Crosstabulation

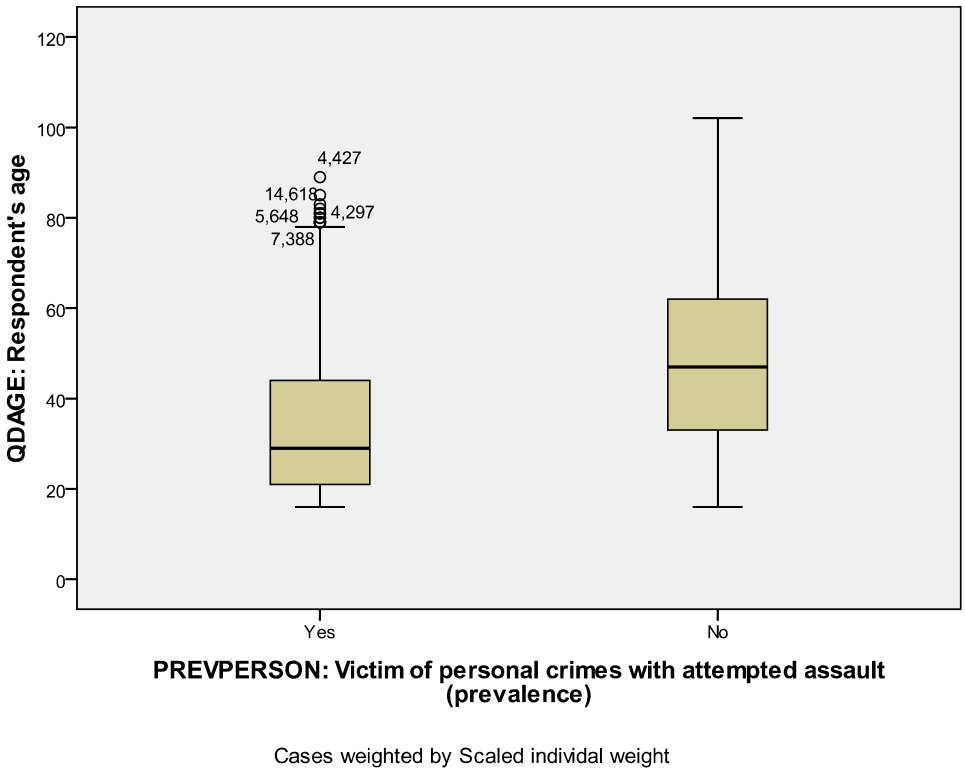
			PREVPERSON: Victim of personal crimes with attempted assault (prevalence)		Total
			1.00 Yes	2.00 No	
QDAGE2: Respondent's age in bands	1 16 - 17	Count	92	376	468
		% within QDAGE2	19.7%	80.3%	100.0%
	2 18 - 19	Count	90	419	509
		% within QDAGE2	17.7%	82.3%	100.0%
	3 20 - 24	Count	190	1157	1347
		% within QDAGE2	14.1%	85.9%	100.0%
	4 25 - 34	Count	212	2172	2384
		% within QDAGE2	8.9%	91.1%	100.0%
	5 35 - 44	Count	147	2733	2880
		% within QDAGE2	5.1%	94.9%	100.0%
	6 45 - 54	Count	127	2665	2792
		% within QDAGE2	4.5%	95.5%	100.0%
	7 55 - 59	Count	40	1183	1223
		% within QDAGE2	3.3%	96.7%	100.0%
	8 60 - 64	Count	29	1145	1174
		% within QDAGE2	2.5%	97.5%	100.0%
	9 65 - 74	Count	21	1721	1742
		% within QDAGE2	1.2%	98.8%	100.0%
	10 75+	Count	12	1466	1478
		% within QDAGE2	.8%	99.2%	100.0%
Total		Count	960	15037	15997
		% within QDAGE2	6%	94.0%	100.0%

(5) This age difference is confirmed by a comparison of the means for victims and non-victims. The table below shows that the mean age for victims of personal crime was 33, which compares to a mean age of 48 for those who were not victims. The median ages are, similarly, very different. A t-test (not shown here) reveals that these mean ages for the two groups are significantly different.

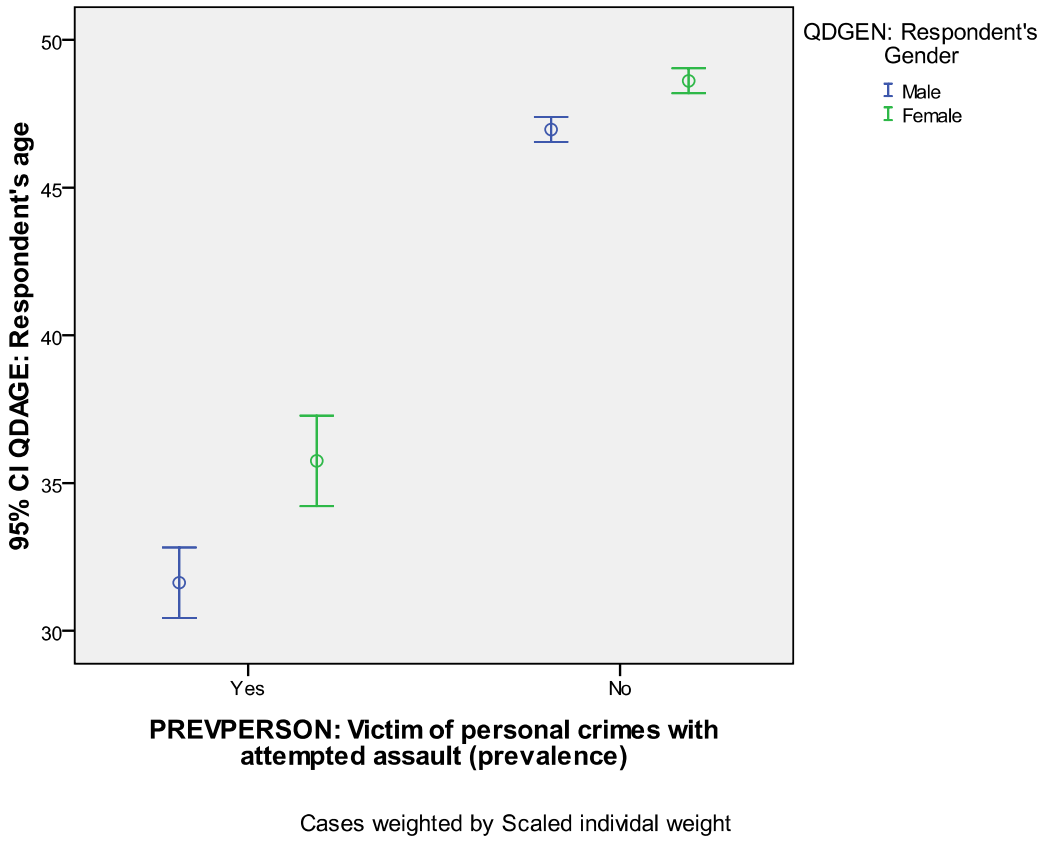
qdage QDAGE: Respondent's age

prevperson PREVPERSON: Victim of personal crimes with attempted assault (prevalence)	Mean	N	Std. Deviation	Median
1.00 Yes	33.33	962	14.976	29.00
2.00 No	47.84	15038	18.692	47.00
Total	46.97	16000	18.808	46.00

(6) The boxplot below compares the median age and interquartile ranges for those who were and were not victims of personal crime. As we might have expected, the range for the bottom quartile of victims is very narrow (i.e. most victims in the bottom quartile are around the same age); whereas, there is a much greater spread in the top quartile (i.e. the ages of the older victims range over a wider number of years). This shows that the age distribution is very unsymmetrical. There are also a number of outliers for the victim boxplot.



(7) The error bar graph below shows that the mean age for male victims of personal crime is lower than the mean age for females. Also, the 95% confidence intervals do not overlap, which means the difference in average age between male and female victims is statistically significant. There is no overlap in the 95% confidence intervals for male and female non-victims, which means that male non-victims are younger on average compared to female non-victims. The difference in mean age for male and female victims is far greater than the difference in mean age for non-victims, however.



3.1.3 To what extent does the Scottish public think that crime is a problem? And are there other issues that are considered more problematic?

Approach/analysis

The SCJS questionnaire begins with some 'warm up' questions about the extent to which respondents consider certain types of social issue a 'problem'. Crime is listed amongst a number of other issues, including unemployment, standards of health care, alcohol abuse, racial discrimination and drug abuse. These questions are each designed to be analysed individually, using simple frequency analysis. However, they can also be analysed together, for example using the 'multiple response' command. This example starts by looking individually at the crime variable (QSPR_4); and then examines this option in the context of the other social problems listed in the survey.

Syntax

(1) You should always start by checking your variable(s) of interest. A quick inspection of QSPR_4 in the 'variable view' window of SPSS reveals that the 'refused' (-2) and 'don't know' (-1) options are not set to missing. You may wish to consider whether it is worthwhile excluding these cases. In this case, the percentage of people who refused or said don't know is very small; therefore, they are excluded here using the following syntax:

```
missing values QSPR_4 (-2,-1).
```

This is an attitudinal question, so the scaled individual weight is applied prior to running the frequency command.:

```
weight by WGTGINDIV_SCALE.
fre  QSPR_4.
```

(2) To identify whether there are other social issues that the public think are more problematic than crime, we could run frequencies for each of the issues and compare the tables. However, it is simpler to run a multiple response command which allows us to analyse the responses to all 10 items together. This can be done via the windows (Analyze > Multiple Response). You start with 'define variables sets' and add each of the variables you want to include in the analysis into the 'variables in set' box. Then indicate that you wish to analyse the counted values for 'a big problem' (which has the value 1). The following syntax is produced, which allows you to explore the proportion of respondents who answered 'a big problem' to each of the 10 items:

```
mult response groups=$SOCPROBS 'Social problems - a big problem'
 (QSPR_1 QSPR_2 QSPR_3 QSPR_4 QSPR_5 QSPR_6 QSPR_7 QSPR_8
QSPR_9 QSPR_21 (1))  /frequencies=$SOCPROBS.
```

Output

(1) The output from SPSS shows that around 54% of the public think that crime is a 'big problem' in Scotland and a further 42% think it is a bit of a problem. You can see in the 'missing' section of the table how small a percentage of people refused or said don't know.

qspr_4 QSPR_4: How much of a problem is CRIME in Scotland today?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00 A big problem	8443	52.8	53.6	53.6
	2.00 A bit of a problem	6602	41.3	41.9	95.4
	3.00 Not a problem	718	4.5	4.6	100.0
	Total	15763	98.5	100.0	
Missing	-2.00 Refused	5	.0		
	-1.00 Don't know	232	1.5		
	Total	237	1.5		
Total		16000	100.0		

(2) The output below shows that the multiple response command presents the results in two ways: Firstly, in terms of the percentage of responses (i.e. of all the answers received across the 10 questions in which the response 'a big problem' was given, what percentage can be attributed to each individual item). This shows that of all the big problems in Scotland, the social problem considered the biggest is drug abuse (21% of all responses), followed by alcohol abuse (19%) and then crime (14%).

Secondly, the output shows the percentage of cases (i.e. the percentage of all those who answered each individual question who noted that this issue was a 'big problem' rather than just a bit of a problem or no problem at all). This confirms that drug abuse was considered the most problematic issue in Scotland today, with 82% of respondents reporting this to be a 'big problem'. This was closely followed by alcohol abuse (77% stating this to be a big problem) and somewhat behind this was crime (57% a big problem).

The result is a little different from the original frequency analysis because some cases with missing data in one or more of the 10 items have been excluded.

To calculate significant differences between responses to different questions, this would have to be done individually using cross-tabulations.

$SOCPROBS Frequencies

		Responses		Percent of Cases
		N	Percent	
$SOCPROBS Social problems - a big problem[a]	qspr_1 QSPR_1: How much of a problem is UNEMPLOYMENT in Scotland today?	6891	11.7%	46.5%
	qspr_2 QSPR_2: How much of a problem is STANDARDS OF HEALTH CARE in Scotland today?	2319	3.9%	15.7%
	qspr_3 QSPR_3: How much of a problem is RACIAL DISCRIMINATION in Scotland today?	2442	4.1%	16.5%
	qspr_4 QSPR_4: How much of a problem is CRIME in Scotland today?	8443	14.3%	57.0%
	qspr_5 QSPR_5: How much of a problem is STANDARDS OF HOUSING in Scotland today?	3503	5.9%	23.7%
	qspr_6 QSPR_6: How much of a problem is DRUG ABUSE in Scotland today?	12175	20.6%	82.2%
	qspr_7 QSPR_7: How much of a problem is STANDARDS OF EDUCATION in Scotland today?	1637	2.8%	11.1%
	qspr_8 QSPR_8: How much of a problem is ALCOHOL ABUSE in Scotland today?	11382	19.3%	76.9%
	qspr_9 QSPR_9: How much of a problem is STANDARDS OF PUBLIC TRANSPORT in Scotland today?	2310	3.9%	15.6%
	qspr_21 QSPR_21: How much of a problem is ANTI SOCIAL BEHAVIOUR in Scotland today?	7979	13.5%	53.9%
Total		59081	100.0%	399.1%

a. Dichotomy group tabulated at value 1.

3.1.4 What attitudes do the Scottish public have towards community sentences?

Approach/analysis

There is a large bank of questions on community sentences in the full sample module, which is asked just after the screener questions for victimisation. Within this bank of questions, the respondents are read out six statements "that other people have made about community sentencing in general" and asked how much they agree or disagree with these statements (a 5 point scale ranging from agree strongly to disagree strongly). The statements are as follows:

- Only criminals who have committed serious crimes should be put in prison
- Prison sentences discourage others from committing crimes
- Prison sentences are a good way to punish offenders for their crimes
- Putting criminals in prison protects the community
- In prison criminals learn new ways to offend
- Prison sentences do not discourage criminals from committing more crimes in the future

These attitudinal questions should, of course, each be analysed individually as a first step within you analysis. However, it is also possible to consider grouping them together into an attitudinal 'scale' which ranges from very negative responses through to very positive ones. This scale can then be used in further analysis.

Syntax

(1) Four of the items are worded in a positive way (QDISATT_1 to QDISATT_4), while the remaining two (QDISSAT_5 and QDISSAT_6) are worded in a negative way. Before they can be grouped together into a scale, all of the items have to be coded in the same way (i.e. running from negative to positive or vice versa). Therefore, you must start by recoding the variables so that they are all ordered in the same way (from negative to positive in this example) and to get rid of the irrelevant responses (refused and don't know). The following syntax will achieve this purpose:

```
recode QDISATT_1 (5=0) (4=1) (3=2) (2=3) (1=4) (else=sysmis) into CSATT1.
recode QDISATT_2 (5=0) (4=1) (3=2) (2=3) (1=4) (else=sysmis) into CSATT2.
recode QDISATT_3 (5=0) (4=1) (3=2) (2=3) (1=4) (else=sysmis) into CSATT3.
recode QDISATT_4 (5=0) (4=1) (3=2) (2=3) (1=4) (else=sysmis) into CSATT4.
recode QDISATT_5 (5=4) (4=3) (3=2) (2=1) (1=0) (else=sysmis) into CSATT5.
recode QDISATT_6 (5=4) (4=3) (3=2) (2=1) (1=0) (else=sysmis) into CSATT6.
exe.
```

You will see here that the values have been changed slightly. Instead of having values 1 to 5, the new variables will have values of 0 to 4. This does not make a great deal of difference, but it means your scale runs from 0 which may makes it

more meaningful to run a 'mean' for example. Note also that the recode has created six new variables, so we have not altered the original variables at all.

(2) Once a set of scale measures running in the same direction have been created, a simple method of testing the reliability of grouping the six measures into one overarching scale exists. This is found in the 'scale' option in SPSS, and performs reliability analysis to determine whether the six items have a close enough correlation that they may be single items that contribute towards an overarching latent variable. The following syntax will achieve this analysis. Note that as this is attitudinal data, the scaled individual weight is used.

```
weight BY WGTGINDIV_SCALE.
reliability
  /variables=CSATT1 CSATT2 CSATT3 CSATT4 CSATT5 CSATT6
  /scale('all variables') all
  /model=alpha
  /statistics=corr
  /summary=total.
```

Output

(2) The reliability analysis produces a standardised Cronbach's Alpha value for the new scale. Cronbach's alpha is a coefficient of reliability (ranging from 0 to 1) that is commonly used to measure internal consistency or reliability of a set of items, for example during psychometric testing. A high value might indicate that the items collectively represent an underlying 'latent trait'. In this case, we have a Cronbach's alpha value of 0.681, which indicates that the variables have reasonably high internal consistency (although many social scientists would prefer this value to be above 0.700). This could be considered just good enough to consider constructing a singular scale (although you should also consider more complex forms of analysis such as factor analysis, principal components analysis or structural equation modelling).

Reliability Statistics

Cronbach's Alpha	N of Items
.681	6

The inter-item correlation matrix shows that none of the item pairings have a particularly high correlation, although this is not really surprising given that the scales have only five points. The fifth and sixth questions appear to be the most highly correlated (with an r value of .599). Those are the questions that asked whether the respondents agreed that "in prison criminals learn more ways to commit crime" and "prisons do not discourage criminals from offending in the future". This indicates that negative responses to community service may cluster more strongly together than positive responses.

Inter-Item Correlation Matrix

	CSATT1	CSATT2	CSATT3	CSATT4	CSATT5	CSATT6
CSATT1	1.000	.212	.241	.227	.191	.223
CSATT2	.212	1.000	.243	.263	.189	.211
CSATT3	.241	.243	1.000	.281	.268	.319
CSATT4	.227	.263	.281	1.000	.269	.296
CSATT5	.191	.189	.268	.269	1.000	.599
CSATT6	.223	.211	.319	.296	.599	1.000

The mean score across the items is shown as 12.6. The way in which the variables were recoded meant that the range ran from 0 to 24, which suggests that this may be a very normally distributed variable (this could be tested using a Kolmogorov-Smirnoff Test or looking at a histogram).

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
12.6223	20.916	4.57345	6

The item-total statistics table shows the Cronbach's alpha value for the scale if each of the items was to be deleted. We can see that removal of any item would make the Cronbach's Alpha value lower than 0.681, so we would not wish to delete any particular item.

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
CSATT1	10.5809	15.442	.331	.671
CSATT2	9.8770	15.821	.340	.664
CSATT3	9.4996	16.529	.420	.640
CSATT4	10.4178	15.080	.413	.639
CSATT5	11.4065	15.270	.475	.619
CSATT6	11.3299	14.574	.516	.603

To create a scale from these items, you simply compute a new scale variable which adds the scores for each variable together. A higher value on the new scale variable would indicate a positive attitude to community sentences, whereas a lower value would indicate a negative attitude.

3.1.5 How common are credit card fraud and identity theft?

Approach/analysis

There is a perception that credit card and identity theft in Scotland are common. Therefore, one quarter sample module (Module C) of the 2008/09 survey asked a set of questions about people's experiences of credit card fraud and identity theft. Here we simply look to see what percentage of survey respondents knew that they had been a victim in terms of: having their credit card and/or bank card used without their permission (CARDVIC_01); having their bank details used without their permission (CARDVIC_02); and having had someone pretend to be the respondent or use their personal details without permission (IDTHEF).

Syntax

In this case, we take the three separate questions above and compute a new variable 'CC_ID_CRIME' which identifies the prevalence of having experienced one of these three types of fraud. The new variable has a value of 1 if the respondent experienced any one of the three crime types, and 0 if they did not. Again, as the analysis is about the experience of individuals, the scaled individual weight is used.

```
compute CC_ID_CRIME=0.
if cardvic_01 = 1  CC_ID_CRIME=1.
if cardvic_02 = 1  CC_ID_CRIME=1.
if idthef =1  CC_ID_CRIME=1.
var lab  CC_ID_CRIME 'Known victim of credit card or identity theft crime'.
value labels  CC_ID_CRIME 1 'Yes' 0 'No'.

weight BY WGTGINDIV_SCALE.
fre  CC_ID_CRIME.
```

Output

The output from the frequency command shown above produces the following table which shows that, despite the perception that this is a common problem, only 1% of the sample as a whole reported that they had been a victim of either credit/bank card fraud or identity theft.

CC_ID_CRIME Known victim of credit card or identity theft crime

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00 No	15833	99.0	99.0	99.0
	1.00 Yes	167	1.0	1.0	100.0
	Total	16000	100.0	100.0	

3.2 Examples using the victim form file (VFF)

3.2.1 What types of crime are most commonly reported to the police?

Approach/analysis

Not all crimes are reported to the police; however, previous crime surveys have shown that some types of crime (especially those that are more serious) tend to be reported more often than others (especially more trivial crimes). Information about whether crimes are reported to the police is contained in the victim form file (VFF). The 2008/09 SCJS VFF contains the variable OFFENCE which represents the final offence coding allocated to each victim form (only one code is applied to each form, so where multiple crimes have occurred within the same offence the most serious coding is applied).

The variable OFFENCE has a total of 58 different values for very specific offences. However, many of these offences are very rare and this results in very small numbers for some offence types. Therefore, for the purposes of most analysis this variable is recoded into crime groups, such as those shown earlier in Table 2.1 of this User Guide. To illustrate this, we have recoded the variable OFFENCE is into a new offence variable OFFTYPE which represents the standard SCJS crime groups.

To examine which of these crime groups are most commonly reported to the police the new variable OFFTYPE can be cross-tabulated against the variable QPOL (which indicates whether or not the police were informed about the incident). For the purposes of presentation, the responses 'refused' and 'don't know' in the QPOL variable have been defined as missing. The incident weight WGTGINC_SCJS is applied since this takes account of series incidents; however, it should be noted that crimes occurring outwith Scotland or outside the 12 month reference period will be excluded from this analysis.

Syntax

To recode OFFENCE into the standard offence groups used in the SCJS:

```
compute OFFTYPE =98.
if (offence=49) OFFTYPE =1.
if (offence=48 or offence=50 or offence=51) OFFTYPE =2.
if (offence=35 or offence=37) OFFTYPE =3.
if (offence=44 or offence=45) OFFTYPE =4.
if (offence=34 or offence=36) OFFTYPE =5.
if (offence=38) OFFTYPE =6.
if (offence=25 or offence=26 or offence=27) OFFTYPE =7.
if (offence=24 or offence=29 or offence=30 or offence=31 or offence=32
or offence=39) OFFTYPE =8.
if (offence=3 or offence=7) OFFTYPE =9.
if (offence=2 or offence=4 or offence=5) OFFTYPE =10.
```

```
if (offence=8 or offence=9 or offence=10 or offence=11 or offence=12 or
offence =13 or offence=14 or offence=15) OFFTYPE =11.
if (offence=19 or offence=20 or offence=21) OFFTYPE =12.
if (offence=17 or offence=18) OFFTYPE =13.
if (offence=41 or offence=46) OFFTYPE =14.
if (offence=55 or offence=56 or offence=57 or offence=58) OFFTYPE =15.
if (offence=1 or offence=6 or offence=16 or offence=22 or offence=23
or offence=28 or offence=33 or offence=40 or offence=42 or offence=43
or offence=47  or offence=52 or offence=53 or offence=54 or offence=59
or offence=60 or offence=61  or offence=62 or offence=63) OFFTYPE =16.

var lab OFFTYPE 'Type of offence recorded for each incident of
victimisation'.
val lab OFFTYPE 1 'mv vandalism' 2 'property vandalism' 3 'theft from
mv' 4 'att theft of/from mv'  5 'theft of motor vehicle' 6 'bicycle theft' 7
'housebreaking' 8 'other household theft'  9 'minor assault' 10 'serious
assault' 11 'sexual assault' 12 'theft from person' 13 'robbery'  14 'other
personal theft' 15 'threats' 16 'other crimes and incidents non-classifiable
or outside the scope of the survey'.
```

To set the 'refused' and 'don't know' responses within QPOL to missing:

```
missing values QPOL (-2,-1).
```

To apply the incidence weight and run a cross-tabulation of OFFTYPE by the policing variable QPOL:

```
weight by WGTGINC_SCJS.
crosstabs
  /tables=OFFTYPE by QPOL
  /format=avalue tables
  /cells=count row
  /count round cell.
```

Output

The table below shows that, overall, the police were informed about just over one third (38%) of all incidents of SCJS crimes reported in the survey. This does vary by offence type, however. The crime types that were most commonly reported were theft of a motor vehicle (92%), serious assault (85%) and housebreaking (64%). On the other hand, crimes such as other household theft (22%) and other personal theft (28%) were far less commonly reported. There are further questions in the victim form that can be used to ascertain why some types of crime are less likely to be reported to the police.

Type of offence recorded for each incident of victimisation * QPOL: Whether police were informed about incident Crosstabulation

			QPOL: Whether police were informed about incident		Total
			Yes	No	
Type of offence recorded for each incident of victimisation	mv vandalism	Count	70784	111169	181953
		% within Type of offence	38.9%	61.1%	100.0%
	property vandalism	Count	68432	97531	165963
		% within Type of offence	41.2%	58.8%	100.0%
	theft from mv	Count	21003	32487	53490
		% within Type of offence	39.3%	60.7%	100.0%
	att theft of/from mv	Count	3643	4998	8641
		% within Type of offence	42.2%	57.8%	100.0%
	theft of motor vehicle	Count	6829	594	7423
		% within Type of offence	92.0%	8.0%	100.0%
	bicycle theft	Count	7904	22637	30541
		% within Type of offence	25.9%	74.1%	100.0%
	housebreaking	Count	16203	9282	25485
		% within Type of offence	63.6%	36.4%	100.0%
	other household theft	Count	30872	110445	141317
		% within Type of offence	21.8%	78.2%	100.0%
	minor assault	Count	106368	163960	270328
		% within Type of offence	39.3%	60.7%	100.0%
	serious assault	Count	21764	3945	25709
		% within Type of offence	84.7%	15.3%	100.0%
	theft from person	Count	9353	10542	19895
		% within Type of offence	47.0%	53.0%	100.0%
	robbery	Count	9457	10240	19697
		% within Type of offence	48.0%	52.0%	100.0%
	other personal theft	Count	24738	65160	89898
		% within Type of offence	27.5%	72.5%	100.0%
Total		Count	397350	642990	1040340
		% within Type of offence	38.2%	61.8%	100.0%

3.2.2 How does the age profile of victims of crime occurring outside pubs/bars/clubs differ from that of other victims?

Approach/analysis

There are concerns in Scotland about the relationship between crime and alcohol. Many incidents of violent crime and anti-social behaviour, for example, occur outside licensed premises such as pubs, bars and clubs. Using data from the crime survey, it is possible to review the age profile of those who are victims of crime in these places. The victim file form collects information about where incidents took place using variables QWH 1-7. Variables QWH1, QWH3 and QWH4 collect information about incidents that occurred in or around the respondent's home. Variables QWH5 and QWH6 collect information about incidents that occurred in or near the respondent's place of work. While, the variable QWH7 identifies various other places where incidents occurred, including 'in/around a pub/bar/nightclub or working men's club'.

Syntax

(1) In order to determine whether the age profile of victims varies depending on where the crimes occur, we need to start by combining the QWH variables into one single variable. Looking at a frequency analysis of QWH1 to QWH7 (not shown here), we can determine that the most common places for crimes to occur were in or outside the respondent's own home, followed by in or near their place of work. Since we may be interested to retain this information, we can compute a new variable OFFWHR with the following categories: in or near home; in or near place of work; in or near pub/club/bar and other.

```
compute OFFWHR =99.
if (QWH1 eq 1) or (QWH1 eq 2)or (QWH3 eq 1) OFFWHR=1.
if (QWH5=1) OFFWHR =2.
if (QWH7 = 1) OFFWHR = 3.
if (QWH7 ge 2) OFFWHR = 4.
var lab OFFWHR 'Where offence took place'.
val lab OFFWHR 1 'In or near home' 2 'In or near workplace' 3 'In or near pub/club/bar' 4 'Elsewhere'.
```

Set the missing values for value 99 of OFFWHR (as these are cases where there is no information on where the offence occurred):

```
missing values OFFWHR (99).
```

Then run a frequency analysis to check the new variable (this is done without weighting to ensure that the distribution of cases across the values of the variable look approximately correct, based on your analysis of the original variables):

```
weight off.
fre OFFWHR.
```

(2) Once this new variable is created, it is a simple matter of examining it against the age profile of victims. However, information about the age of the respondent is not contained in the victim form file – it can be found instead in the respondent file. This requires merging the variables QDAGE and QDAGE2 from the RF to the VFF (for instructions on how to do this, see section 3.4.1 of this User Guide).

One simple way of exploring age difference is to run a cross-tabulation between OFFWHR and QDAGE2 to see whether the age bands differ for each crime location. A chi-square test can be used to determine if the differences are statistically significant. To do this, use the following syntax:

```
weight by WGTGINC_SCJS.

crosstabs
  /tables=OFFWHR by QDAGE2
  /format=avalue tables
  /cells=count row
  /count round cell.
```

(3) Another approach is to explore the mean age of victims according to where they were victimised. Again, we use the scaled incidence weight to reflect the fact that some victim forms are counting more than one incident. To compare mean age of victims for incidents occurring in different locations, you can use the following syntax:

```
means tables= QDAGE by OFFWHR
  /cells mean count stddev median
  /statistics anova.
```

Output

(1) The frequency table, below, allows us to check the new variable OFFWHR to determine how the cases are distributed across the new categories.

OFFWHR Where offence took place

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00 In or near home	3252	56.3	66.2	66.2
	2.00 In or near workplace	488	8.5	9.9	76.1
	3.00 In or near pub/club/bar	256	4.4	5.2	81.3
	4.00 Elsewhere	920	15.9	18.7	100.0
	Total	4916	85.2	100.0	
Missing	99.00	856	14.8		
Total		5772	100.0		

The majority of incidents occurred in or around the victim's own home, while only a relatively small proportion occurred in or near a pub, club or bar. However, weighted analysis using WGTGINC_SCJS would be required in order to report the correct percentages, as this would take account of the series incidents.

(2) When OFFWHR is cross-tabulated against QDAGE2, the results show that incidents of crime occurring near pubs, clubs or bars are more likely to involve younger victims compared to those that occur at home or in the workplace. Incidents of crime occurring 'elsewhere' also tend to involve young victims.

qdage2 QDAGE2: Respondent's age in bands * OFFWHR Where offence took place Crosstabulation

			OFFWHR Where offence took place				
			1.00 In or near home	2.00 In or near workplace	3.00 In or near pub/ club/bar	4.00 Elsewhere	Total
QDAGE2: Respondent's age in bands	1 16 - 17	Count	11674	2412	4062	38499	56647
		% within QDAGE2	1.9%	2.0%	5.0%	18.6%	5.4%
	2 18 - 19	Count	17978	2192	16289	22903	59362
		% within QDAGE2	2.9%	1.8%	20.1%	11.0%	5.7%
	3 20 - 24	Count	60730	16263	30046	32529	139568
		% within QDAGE2	9.7%	13.2%	37.1%	15.7%	13.4%
	4 25 - 34	Count	133311	28028	16167	37813	215319
		% within QDAGE2	21.2%	22.8%	19.9%	18.2%	20.7%
	5 35 - 44	Count	159713	24419	7230	29377	220739
		% within QDAGE2	25.4%	19.9%	8.9%	14.2%	21.2%
	6 45 - 54	Count	104709	32657	5553	23107	166026
		% within QDAGE2	16.7%	26.6%	6.8%	11.1%	16.0%
	7 55 - 59	Count	46008	10328	1052	8040	65428
		% within QDAGE2	7.3%	8.4%	1.3%	3.9%	6.3%
	8 60 - 64	Count	40736	4349	345	4763	50193
		% within QDAGE2	6.5%	3.5%	.4%	2.3%	4.8%
	9 65 - 74	Count	37039	1923	338	7008	46308
		% within QDAGE2	5.9%	1.6%	.4%	3.4%	4.5%
	10 75+	Count	16926	175	0	3319	20420
		% within QDAGE2	2.7%	.1%	.0%	1.6%	2.0%
Total		Count	628824	122746	81082	207358	1040010
		% within QDAGE2	100.0%	100.0%	100.0%	100.0%	100.0%

(3) Finally, the output below shows that there is a significant difference in the mean age of victims of crime for incidents occurring in different locations. Where crimes occurred in or near the home or workplace, victims tended to be aged on average around 40. However, crimes that occurred in or near pubs, clubs or bars tended to involve victims who were on average in their mid twenties. Interestingly, although many of the crimes occurring elsewhere involved very young victims, the overall average age is 32. In other words, the age profile of victims of crimes occurring outside pubs, bars and clubs is generally younger than for victims of crimes that occur elsewhere.

Report

qdage QDAGE: Respondent's age

OFFWHR Where offence took place	Mean	N	Std. Deviation	Median
1.00 In or near home	41.92	628825	15.432	40.00
2.00 In or near workplace	39.36	122747	13.001	41.00
3.00 In or near pub/club/bar	26.51	81081	10.277	23.00
4.00 Elsewhere	32.06	207357	15.987	27.00
Total	38.45	1040011	15.803	37.00

3.3 Examples using the self-completion data file (SCF)

3.3.1 Is partner abuse more prevalent amongst men or women?

Approach/ analysis

An extensive set of questions are contained in the self-completion component of the 2008/09 SCJS. Rather than asking respondents directly if they have experienced partner abuse, however, respondents are asked whether they have experienced any of a series of acts or behaviours by a partner since they were age 16. Because the information we need to determine simply whether any respondent has ever been a victim of partner abuse is contained in multiple questions, it is necessary to create a new derived variable (PABUSE).

Once the new variable has been created, this can be cross-tabulated against the gender variable (QDGEN) to show the proportion of men and women reporting partner abuse experience. (Unlike the victim form file, the self-completion data file already contains demographic variables such as age and sex). To test whether or not any gender difference is statistically significant, a chi square test will be applied. The scaled individual weight WGTGINDIV_SC_SCALE will be used here since we are looking at the characteristics of individuals, rather than households.

Syntax

(1) To create the derived variable PABUSE from the individual questions contained in the data file:

```
compute PABUSE=99.
if (da_1i_01=1 or da_1i_02=1 or  da_1i_03=1 or da_1i_04=1 or da_1i_04=1
or da_1i_05=1 or da_1i_06=1 or da_1i_07=1 or da_1i_08=1 or da_1i_09=1
or da_1i_10=1 or da_1i_11=1 or da_1i_12=1 or da_1iii_01=1 or
da_1iii_02=1 or da_1iii_03=1 or da_1iii_04=1 or da_1iii_05=1 or
da_1iii_06=1 or da_1iii_07=1) PABUSE=1.
if (da_1i_01=0 and da_1i_02=0 and  da_1i_03=0 and da_1i_04=0 and
da_1i_04=0 and da_1i_05=0 and da_1i_06=0 and da_1i_07=0 and
da_1i_08=0 and da_1i_09=0
and da_1i_10=0 and da_1i_11=0 and da_1i_12=0 and da_1iii_01=0 and
da_1iii_02=0 and da_1iii_03=0 and da_1iii_04=0 and da_1iii_05=0 and
da_1iii_06=0 and da_1iii_07=0) PABUSE=0.

var lab PABUSE 'ever experienced partner abuse'.
val lab PABUSE 1 'yes' 0 'no' 99 'missing'.
missing values PABUSE (99).
```

To apply the individual weight and run a frequency of the derived variable PABUSE to identify overall prevalence of partner abuse:

```
weight by WGTGINDIV_SC_SCALE.
frequencies variables=PABUSE
  /order=analysis.
```

(2) To run a cross-tabulation of PABUSE and QDGEN with a chi-square test:

```
crosstabs
  /tables=PABUSE by QDGEN
  /format=avalue tables
  /statistics=chisq
  /cells=count column
  /count round cell.
```

Output

(1) The frequency table below shows that 18% of respondents to the self-completion questionnaire reported that they had been a victim of partner abuse at some point since the age of 16.

PABUSE ever experienced partner abuse

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00 no	8223	74.9	81.8	81.8
	1.00 yes	1836	16.7	18.2	100.0
	Total	10059	91.7	100.0	
Missing	99.00 missing	915	8.3		
Total		10974	100.0		

(2) The cross-tabulation of PABUSE by QDGEN indicates that there was a significant difference between men and women in terms of their likelihood of being a victim of partner abuse. Women were more likely to report being victims of partner abuse than men, with 21% of female respondents compared to 15% of male respondents having done so. The chi-square test indicates that this result is statistically significant ($p<.001$) and, therefore, we can confidently report that partner abuse is more prevalent amongst women than men. However, in percentage terms, the difference is not great so we would be equally interested in finding out more about partner abuse amongst both male and female victims.

PABUSE ever experienced partner abuse * qdgen QDGEN: Respondent's Gender Crosstabulation

			qdgen QDGEN: Respondent's Gender		Total
			1.00 Male	2.00 Female	
PABUSE ever experienced partner abuse	.00 no	Count	4043	4180	8223
		% within QDGEN	84.7%	79.1%	81.8%
	1.00 yes	Count	732	1103	1835
		% within QDGEN	15.3%	20.9%	18.2%
Total		Count	4775	5283	10058
		% within QDGEN	100.0%	100.0%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	51.766[a]	1	.000		
Continuity Correction[b]	51.394	1	.000		
Likelihood Ratio	52.142	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	51.761	1	.000		
N of Valid Cases	10058				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 871.16.

b. Computed only for a 2x2 table

3.3.2 What percentage of the Scottish population has taken drugs, and does this vary by age and sex?

Approach/ analysis

The SCJS provides the only national prevalence data on drug use. In 2008/09, the self completion questionnaire contained a series of questions in which respondents were asked to identify which items from a list of substances they had ever taken, which they had taken within the last 12 months, and which they had taken within the last month. For the purposes of this example, the 'ever' question will be used. Thus, the results will not show current drug use, but rather lifetime prevalence of drug use.

In order to undertake this analysis, a similar process to that in the previous example of partner abuse is required to create a new derived variable 'EVERDRUG' to measure lifetime drug use. A frequency analysis of this derived variable will determine overall prevalence of drug use, and a cross-tabulation by age and sex will determine the extent to which drug use varies for different demographic groups. This difference will be illustrated using an error bar graph. Once again, the individual weight WGTGINDIV_SC_SCALE will be used as we are looking at individual differences.

Syntax

(1) To create the derived variable EVERDRUG:

```
compute EVERDRUG=99.
if (qeve_01=1 or qeve_02=1 or qeve_03=1 or qeve_04=1 or qeve_05=1 or
qeve_06=1 or qeve_07=1 or qeve_08=1 or qeve_09=1 or qeve_11=1 or
qeve_12=1  or qeve_13=1 or qeve_14=1 or qeve_15=1 or qeve_16=1 or
qeve_17=1) EVERDRUG=1.
If (qeve_01=2 and qeve_02=2 and qeve_03=2 and qeve_04=2 and
qeve_05=2 and qeve_06=2 and qeve_07=2 and qeve_08=2 and qeve_09=2
and qeve_11=2 and qeve_12=2
and qeve_13=2 and qeve_14=2 and qeve_15=2 and qeve_16=2 and
qeve_17=2) EVERDRUG=0.
var lab EVERDRUG 'Whether respondent has ever taken drugs'.
val lab EVERDRUG 1 'yes' 0 'no' 99 'missing'.
missing values EVERDRUG (99).
```

To apply the individual weight and run a frequency analysis of EVERDRUG:

```
weight by WGTGINDIV_SC_SCALE.

frequencies variables=EVERDRUG
  /order=analysis.
```

(2) To analyse lifetime drug use by sex:

```
crosstabs
  /tables=QDGEN by EVERDRUG
  /format=avalue tables
  /statistics=chisq
  /cells=count row
  /count round cell.
```

(3) To analyse lifetime drug use by age (recoding the QDAGE2 variable to group the age 16-24 year olds into one band and those aged 65 or over into one band):

```
recode QDAGE2 (1 thru 3=1) (4=2) (5=3) (6=4) (7=5) (8 thru highest=6)
(else=sysmis) into QDAGE3.
var lab QDAGE3 'Recode of QDAGE2 age bands'.
val lab QDAGE3 1 '16-24 years' 2 '25-34 years' 3 '35-44 years' 4 '45-54
years' 5 '55-64 years' 6 '65+ years'.
```

(4) To create an error bar graph showing age and sex differences in lifetime drug use:

```
graph
  /errorbar(ci 95)= QDAGE by EVERDRUG by QDGEN.
```

Output

(1) The frequency table below shows that around a quarter (26%) of those who responded to the self-completion questionnaire indicated that they had taken an illegal drug at least once during their lifetime.

EVERDRUG Whether respondent has ever taken drugs

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00 no	8142	74.2	74.4	74.4
	1.00 yes	2809	25.6	25.6	100.0
	Total	10950	99.8	100.0	
Missing	99.00 missing	24	.2		
Total		10974	100.0		

(2) Looking at the row percentages in the table below, you can see that prevalence of drug use amongst men (31%) was significantly higher than for women (20%). The chi-square test result indicated that this difference was statistically significant ($p<.001$), although the test results are not shown here.

qdgen QDGEN: Respondent's Gender * EVERDRUG Whether respondent has ever taken drugs

Crosstabulation

			EVERDRUG Whether respondent has ever taken drugs		Total
			.00 no	1.00 yes	
qdgen QDGEN: Respondent's Gender	1.00 Male	Count	3590	1640	5230
		% within QDGEN	68.6%	31.4%	100.0%
	2.00 Female	Count	4552	1169	5721
		% within QDGEN	79.6%	20.4%	100.0%
Total		Count	8142	2809	10951
		% within QDGEN	74.3%	25.7%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	170.967[a]	1	.000		
Continuity Correction[b]	170.395	1	.000		
Likelihood Ratio	171.256	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	170.952	1	.000		
N of Valid Cases	10951				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 1341.53.

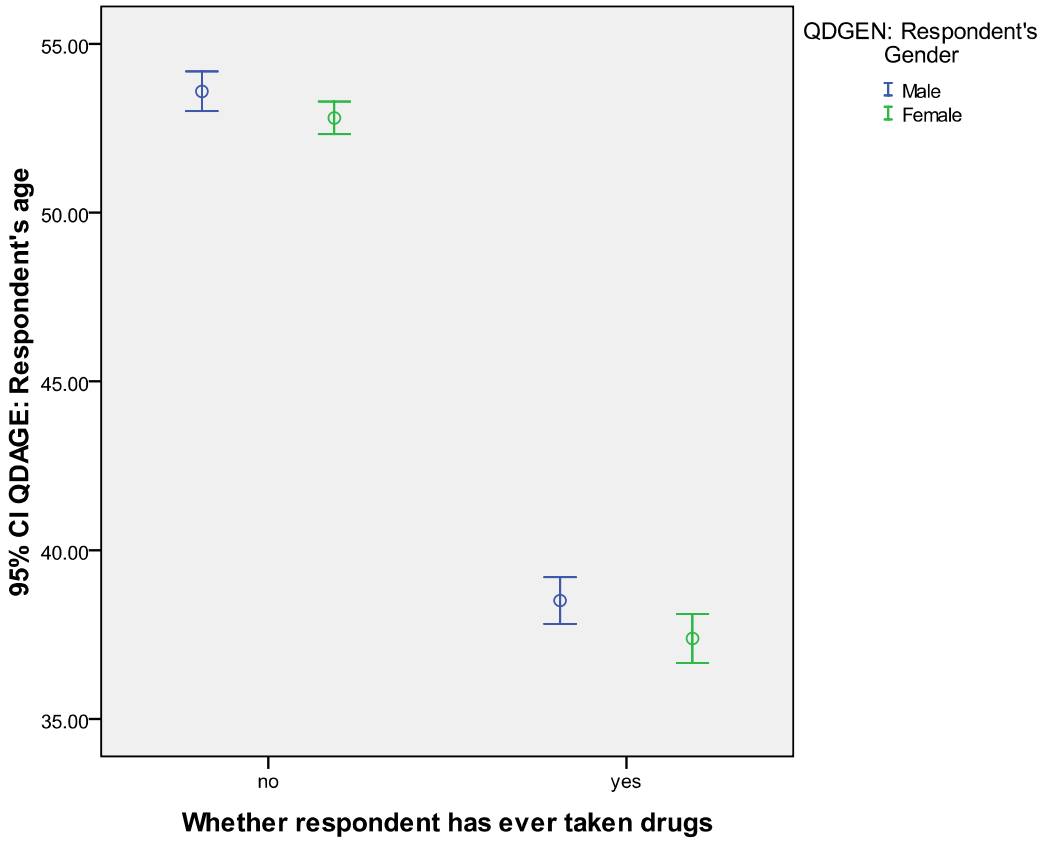
b. Computed only for a 2x2 table

(3) Using the recoded age variable, the cross-tabulation below shows that lifetime prevalence of drug use was greatest amongst those in the youngest age bands – especially those aged 25 to 34 amongst whom about half (49%) had used an illegal drug at some time. Prevalence reduced with age, with only one in 20 of those aged 65 or over stating that they had ever taken an illegal drug.

QDAGE3 Recode of QDAGE2 age bands * EVERDRUG Whether respondent has ever taken drugs Crosstabulation

			EVERDRUG Whether respondent has ever taken drugs		Total
			.00 no	1.00 yes	
QDAGE3 Recode of QDAGE2 age bands	1.00 16-24 years	Count	933	658	1591
		% within QDAGE3	58.6%	41.4%	100.0%
	2.00 25-34 years	Count	835	793	1628
		% within QDAGE3	51.3%	48.7%	100.0%
	3.00 35-44 years	Count	1311	658	1969
		% within QDAGE3	66.6%	33.4%	100.0%
	4.00 45-54 years	Count	1501	413	1914
		% within QDAGE3	78.4%	21.6%	100.0%
	5.00 55-64 years	Count	692	145	837
		% within QDAGE3	82.7%	17.3%	100.0%
	6.00 65+ years	Count	2871	142	3013
		% within QDAGE3	95.3%	4.7%	100.0%
Total		Count	8143	2809	10952
		% within QDAGE3	74.4%	25.6%	100.0%

(4) An error bar graph (shown below) confirms that there is a highly significant difference in lifetime prevalence of drug use by age (with the mean age of drug users being around 35 and the mean age of non-drug users being around 50). Comparing the error bars for males and females indicates that the mean age for male drug users is slightly higher than for female drug users. However, the confidence intervals overlap, so the difference is not statistically significant. In other words, the average age for female drug users is around the same as for male drug users.


55.00
50.00
45.00
40.00
35.00
95% CI QDAGE: Respondent's age
no
yes
Whether respondent has ever taken drugs
QDGEN: Respondent's Gender
Male
Female

3.4 Examples that involve merging datasets

While there are already some shared derived variables contained in the SCJS 2008/09 data files, data users who have created further derived variables may wish to add these variables to one of the other data files. This will involve merging datasets. This is complicated to an extent because the victim form file (VFF) has a different unit of analysis (i.e. crime incidents) to the respondent file (RF) and the self-completion file (SCF) (i.e. respondents). The first two examples below take you through how to merge data from the respondent data file to the victim form file, and vice versa. The final example shows how to merge data from the self-completion data file to the respondent data file.

3.4.1 Merging variables from the RF to the VFF

Users who wish to use incidents of crime (rather than individuals) as their unit of analysis may wish to add information about the victim that is only held in the respondent data file, e.g. attitudinal or socio-demographic data. This process involves taking information held for one case only in the RF and merging it on to numerous cases that share the same unique identifier in the VFF. This type of 'one to many' merging can be achieved in two ways in SPSS, one of which involves using the command windows and the other is purely syntax driven.

File merging example 1: the command windows method

The first method of file merging involves using the SPSS command windows. It is useful to learn how to merge files using this method, as it can be a quick and simple alternative to the syntax method.

1. Open both the VFF and the RF data files in SPSS.

2. Sort both files ascending on the variable SERIAL using the following syntax:

   ```
   sort cases by SERIAL(A).
   ```

3. Ensure that the VFF is the 'active dataset' (i.e. the one you have open in front of you).

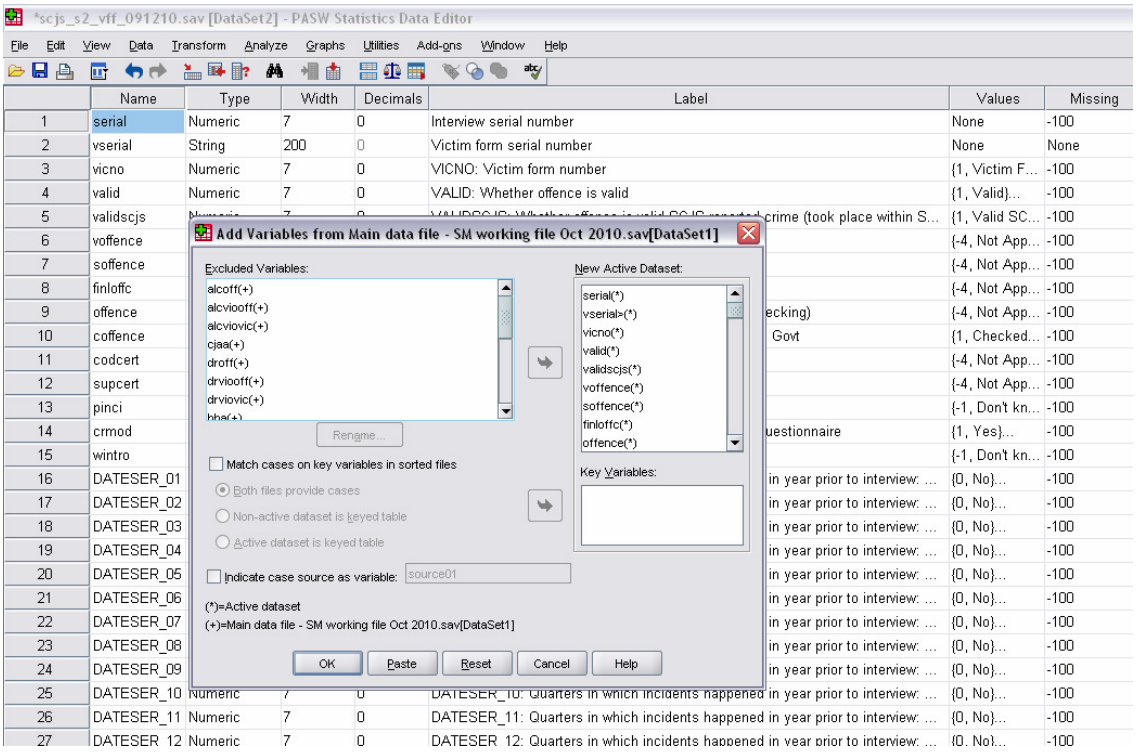
4. In the main menu, go to 'Data', click on 'Merge files', and then 'Add variables'.

5. Click on 'An active dataset' and select the RF from the list. Press continue.

6. A new command window, 'Add variables from…', will open (see Figure 3.1). The dialogue box 'new active dataset' contains the names of all the variables in the primary dataset (VFF) and all those from the secondary dataset (RF). SPSS automatically assumes you wish to merge the files

fully together and so the 'new active dataset' box contains every variable from both files combined. Those marked with an * indicate variables from the primary dataset (VFF), while those marked with a + indicate those that will be merged in from the secondary data file (RF).

Figure 3.1: Screenshot of the Merge Files dialogue box in SPSS



The 'Excluded variables' box at this stage contains only those variables from the secondary dataset that are already also contained in the primary dataset, so there is no need to merge these on. If you do not wish all the variables from the RF to be merged onto the VFF, then you need to move those that should not be merged back from the 'new active dataset' to the 'excluded variables' box. This involves going down through the list of + variables and clicking on the upper, left pointing arrow button. You do not need to do this one by one – you can select a whole list of variables to transfer by clicking on the first one in the list, holding down the shift key and then clicking on the last one in the list.

(Sometimes, if you only want to add one or two variables to the new dataset, it is easier to save a small data file containing only those variables you want to transfer, and then using this new small data file to do the merging process).

7. Once you have all the variables that you want to transfer in the 'new active dataset' dialogue box, select 'match cases on key variable in sorted files'.

8. In the 'Excluded variables' dialogue box find SERIAL(+) from the excluded variables list and press the lower, right-pointing arrow to place this variable in the 'key variables' dialogue box. This allows you to match the files on the basis of the individual's unique identifier.

9. Click on 'non-active dataset is keyed table' which tells SPSS that each of the cases in the VFF must be matched with a value from the RF. Press OK.

10. As a matter of default when you run this analysis a message will appear stating 'Warning: Key match will fail if data are not sort in ascending order of key variables'. (This is to remind you that both files need to be sorted in the same order by the same variable).

11. Press OK again and (provided the data files are sorted in the same order) SPSS will merge the variables from RF onto VFF, matching one to many.

In this case, respondents who have completed more than one victim form (and so have more than one record in the victim file) will have the same information transferred to each record (because the files are matched on serial number). For those respondents with no experience of victimisation, no information will be transferred over to the victim file.

File merging example 2: the syntax method

There is an alternative method to match variables from one file to another which involves using the 'match files' command. This method is syntax, rather than windows, driven and involves reading the files into SPSS (using the 'data list' or 'get data' commands), defining dataset names for each file (using the 'dataset name' command), and then using 'match files' to add the cases from each file.

As an example, you may wish to merge the variable QDGEN from the respondent file to the victim form file (in order to analyse the gender of victims of different types of crime incidents). Start by using the 'get file' command to open the respondent data file. Ensure cases are sorted by the unique identifier, SERIAL, and that any weights are turned off (note that if there is a weight on during the match file process this weight will also be transferred to the victim form file.

```
get file = 'insert your file pathway for the respondent file here'.

Sort cases by SERIAL(A).

weight off.
```

Save a new outfile that contains only the unique identifier, SERIAL (used for matching cases), and any variables that you wish to merge on to the victim form file (in this example, QDGEN). This involves using the 'keep' command.

```
save outfile = 'insert file pathway for new data file X here'
 /keep SERIAL QDGEN.
```

Use the 'get file' syntax command to get the data file that you wish to merge your variables on to, in this case the victim form file. Again, ensure that the cases are sorted on the key matching variable, SERIAL, or the matching will fail. Note that the key matching variable can occur more than once in the 'receiving' file but it must appear only once in the 'donating' file.

```
get file = 'insert your file pathway for the victim form file here'.

sort cases by SERIAL(A).
```

The final stage is to use the 'match files' command to merge the variables from your new data file (created at the previous step) into the victim form file (now open as the active file) based on the matching variable SERIAL.

```
match files file = *
 /table = 'insert file pathway for new data file X here'
 /by SERIAL.

execute.
```

The variable QDGEN will now appear at the end of the victim form file. It is important when reporting analysis from the VFF data that you remember the unit of analysis is crime incidents and NOT individuals, so that you do not present a misleading picture of crime victims. So in this example which involves merging the variable QDGEN from the RF to the VFF, if you were to run a cross-tabulation of QDGEN against the VFF variable PINCI (whether the victim form represented a single crime incident or a series of related incidents), you would get the output shown in Figure 3.2. To interpret the data, you would state that 47.6% of single incidents were reported by male victims compared with 42.7% of series incidents. In other words, series incidents were more likely to be reported by female victims. It would NOT be accurate to say that 47.6% of male victims reported single incidents because the victim is not the unit of analysis here. As a rule of thumb, the category of interest (reflecting the unit of analysis) should be in the columns rather than the rows.

Figure 3.2: SPSS output showing single and series incidents by gender

Respondent gender by single or series incident

		PINCI: Incident type (single or series)		Table Total
		1 Single	2 Series	.00
		Col %	Col %	Col %
QDGEN: Respondent's Gender	1 Male	47.6%	42.7%	46.5%
	2 Female	52.4%	57.3%	53.5%
Table Total		100.0%	100.0%	100.0%

3.4.2 Merging variables from the VFF to the RF

Data users may also wish to take information about incidents of crime victimisation and aggregate this to individual level, so that it can be analysed alongside other data from the RF and SCF. As noted earlier, some derived victimisation variables from the VFF have already been appended to the individual level data files, and data users should ensure that the variables they wish to analyse are not already there. However, if the crime variables in the RF do not meet the user's needs (for example, you wish to include 'non-valid' cases or consider alternative crime categories) it is necessary to merge variables from the VFF to the main data file. The process described in section 3.4.1 is not appropriate for merging the files here, because there is potentially more than one piece of information that could be appended to each case in the RF (since each individual may have completed more than one victim form). In other words, many to one merging is required. This is a slightly more complex process which requires that the cases in the VFF be 'aggregated' to individual level before the variables can be merged onto the RF.

Two examples of merging variables from the VFF to the RF are provided here. The first, simpler, example explains how to merge information about the reporting of crimes to the police (using variable QPOL) from the VFF to the RF. The second, more complex, example explains how to create prevalence and incidence variables for a type of victimisation (VANDALISM) and aggregate this from the VFF to the RF.

File merging example 3: aggregating simple variables

The RF contains derived variables indicating whether or not the respondent was a victim of crime, and what type of crime they were victims of, but it does not contain any further information about the victimisation experience. Therefore, users may wish to add additional information to the RF, such as how many respondents had contact with the police as a result of being a victim of crime during the last 12 months. The variable QPOL in the VFF indicates whether the police were informed about each incident (or series of incidents) of crime. It is coded 1 for 'yes' and 2 for 'no'. This example explains how to aggregate the multiple records for this variable and merge the new variable onto the RF (so that this information is recorded at the individual level).

There are two stages to this process. Stage one is to 'flatten' or 'aggregate' data from the victim form file using the 'Aggregate Data' command which can be found under the 'Data' menu in SPSS. This command aggregates groups of cases in the active dataset (the VFF) into single cases. The steps for aggregating the variable QPOL are described in detail below, including all necessary syntax. The syntax can be varied, as required, depending on the needs of the user. Stage two is to attach the aggregated data to the main data file (this has already been described in section 3.4.1).

Stage 1: To flatten the data

1. Start by opening the victim file (SCJS_S2_VFF_091210) and turn off all weights and filters.

```
weight off.
filter off.
use all.
execute .
```

2. Run a frequency analysis of the variable QPOL. This needs to be recoded for two reasons: firstly, the raw variable includes values -1 for don't know and -2 for refused, which are irrelevant here so these should be set to missing; and secondly, we ideally want any cases where the police were not informed to have a value of 0 rather than 2 (this will be explained further at step 6, below). This can be done using the following syntax:

```
fre QPOL.

recode QPOL (1=copy) (2=0) (else=sysmis) into QPOL2.
var lab QPOL2 'Whether police were informed about incident - recoded'.
val lab QPOL2 1 'yes' 0 'no'.
```

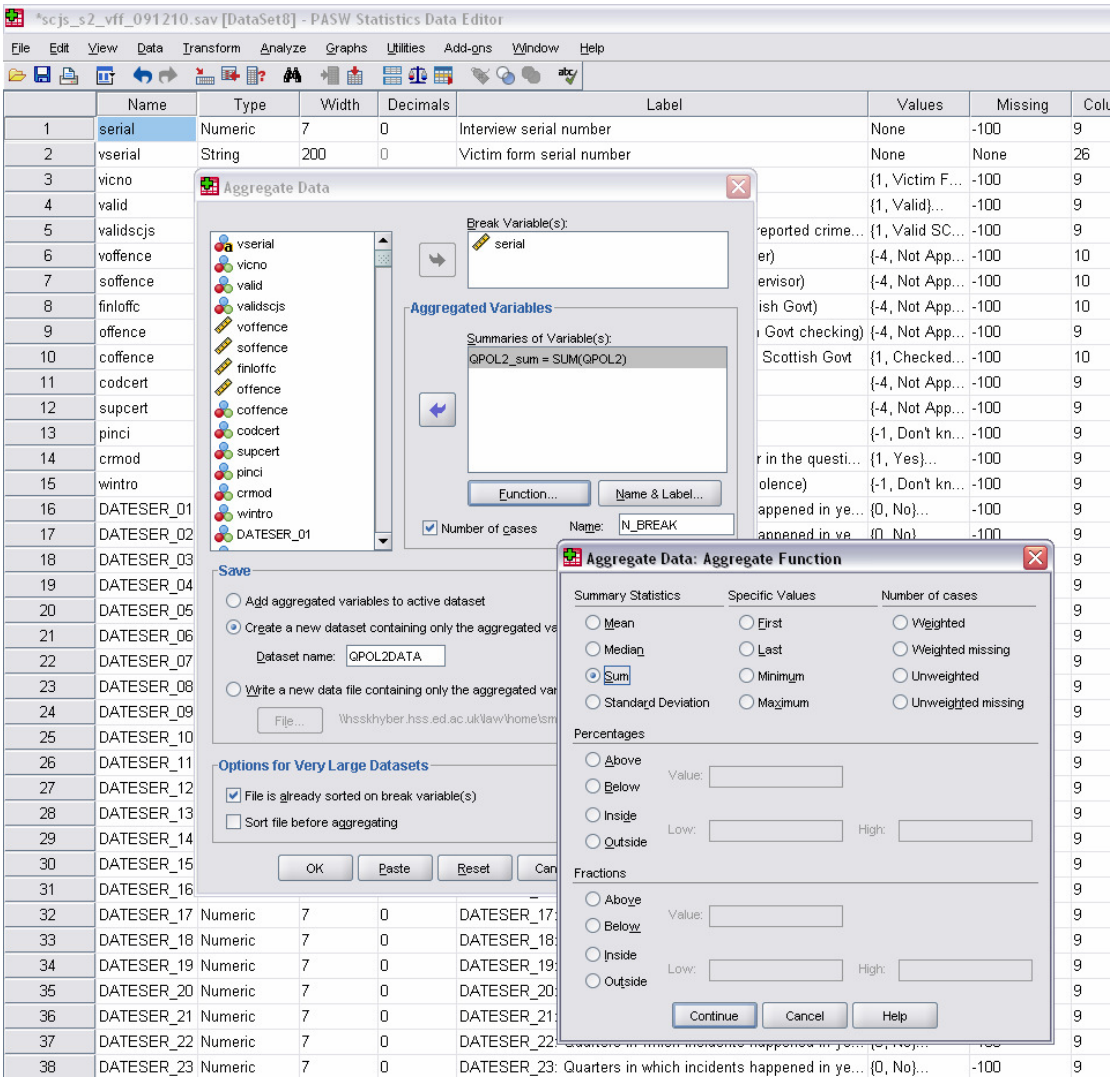
   Note that in addition to the 'refused' and 'don't know' responses, some interviews were terminated prior to completing the full victim form questionnaire (e.g. if they were out of scope of the survey) so these will also be included in the missing cases.

3. Open menu option 'Data' and click on 'Aggregate'. This brings up the 'Aggregate Data' dialogue box, as shown in Figure 3.3 below.

4. The long box on the left contains all the variables in the VFF. The box on the top right titled 'break variable(s)' allows you to indicate how you wish the data to be grouped. Click on the variable SERIAL in the list of variables on the left and transfer it to the 'break variable(s)' box using the right-pointing arrow. This will create one record for each respondent who was a victim of at least one incident of crime.

5. The 'aggregated variables' box allows you to add all those variables that you wish to aggregate using some form of summary function. Find QPOL2 in the list of variables on the left and transfer it to the 'summaries of variables' box by clicking on the right-pointing arrow. You will see this has been done in the screenshot above.

Figure 3.3: Screenshot of the Aggregate Data dialogue box in SPSS



6. Once QPOL2 has been transferred over to the 'summaries of variables' box, click on 'function'. A new dialogue box called 'Aggregate Data: Aggregate Function' opens which allows you to specify a function to use to calculate the aggregated data values for QPOL2. The aggregate functions for numeric variables include mean, median, standard deviation and sum. You can also summarise on the basis of specific values, number of cases (weighted or unweighted) or using percentages or fractions.[6]

   In this case, for each individual we simply want to count the number of incidents of victimisation in which the police were contacted following a crime (i.e. where QPOL = 1). Therefore, under the 'Summary Statistics' list click on the option 'sum'. This is the reason that the variable QPOL was recoded at step 2, so that incidents involving no police contact have a value of zero.

[6] For further information on using the 'aggregate data' command, refer to the online help manual in SPSS.

7. You can add a variable name and label for the new aggregate variable by clicking on the 'Name & Label' dialogue box. And, if you want to create another variable that indicates how many cases were aggregated for each individual (i.e. how many victims forms they completed), tick the box marked 'Number of cases'.

8. In the 'Save' box there are three options for saving the new aggregated variables. It can be appended to the active dataset (which means that each case in the dataset with the same serial number will have the same value). Alternatively, there are two ways of saving the data to a new data file: by creating a new dataset (i.e. a new dataset is opened on screen, but not saved to a directory) or by writing to a new data file (i.e. a new dataset is saved to a directory, but not opened on screen). In the screenshot shown here, a new dataset with the label 'QPOL2DATA' is created.

9. Finally, under the box 'Options for Very Large Datasets' you can indicate whether the data has already been sorted on the break variable(s) or whether sorting is required prior to aggregating. In this case, the data file is already sorted by the break variable SERIAL.

10. If you click on the 'paste' key, the following syntax is created:

```
DATASET DECLARE QPOL2DATA.
AGGREGATE
 /OUTFILE='QPOL2DATA'
 /PRESORTED
 /BREAK=serial
 /QPOL2_sum=SUM(QPOL2)
 /N_BREAK=N.
```

11. After running the above syntax a new untitled dataset is opened on screen (this file is not saved to a directory, so closing the file loses the data). Click on the new dataset, and it should contain three variables:
    a. The first variable is SERIAL, and each case now has a unique value representing a different respondent (i.e. the data has been aggregated from crime incident to individual level).
    b. The second variable is QPOL2_SUM, which will have a value between 0 and 5 (where 0 represents a victim of crime where no cases were reported to the police and 5 represents a victim of five separate crime incidents all of which were reported to the police (any cases with missing data in the VFF will be contained in the new dataset but no value will be assigned to this variable).
    c. The third variable is and N_BREAK which indicates how many cases have been aggregated for each respondent. This is useful information, because from this we can determine whether all, some or none of the crime incidents experienced by repeated victims were brought to the attention of the police depending on how many incidents of victimisation they experienced.

Stage 2: To merge the new aggregated variables onto the RF data file:

This involves using the procedures described in section 3.4.1. In this case, the RF would be the 'active' dataset open in front of you and the dataset you select for merging variables from would be the new untitled dataset labelled QPOL2DATA. Match cases using the variable SERIAL and merge the new variables onto the RF data file. Ensure that 'non-active dataset is keyed table' is highlighted so as to ensure that all of the RF cases receive a value from the new dataset. You should now have two new variables added to your RF data file: QPOL2_SUM and N_BREAK. This will allow you to analyse these data at the level of the individual.

File merging example 4: aggregating complex variables

The second example is slightly more complex, as it involves taking one variable in the VFF and creating multiple new variables in the RF. This process involves the use of vector loops, rather than straightforward data aggregation; however, many of the principles are the same. This example shows users how to recreate some of the prevalence and incidence variables for crime victimisation that are contained in the RF. As stated previously, the RF dataset already includes numerous derived variables for crime victimisation; however, data users may wish to construct reconfigured victimisation variables if the existing crime groupings contained in the SCJS dataset are too narrow or too wide for their purposes. The example given here shows how to recreate variables for three broad crime categories (VANDALISM, AQUISITIVE CRIME AND VIOLENT CRIME) and aggregate these from the VFF to the RF; however, the steps could be adapted to create any other crime groupings, or for any other variables, you wish to use.

When data are aggregated from the VFF to the RF, there are a number of stages to follow in order to create valid SCJS data. These four stages are detailed below in order that users could (if they wished) replicate the creation of SCJS offence variables or choose to create different variables, depending on their needs:

- Stage 1 involves excluding certain crime incidents from the aggregation process. This is done in the SCJS because some cases are not comparable with police recorded crime data. The cases that are considered 'valid' for SCJS purposes must fulfil three criteria:
    - They must have occurred within Scotland.
    - They must have occurred within the 12 months prior to the month of interview.
    - They must be crimes that are considered classifiable for the purposes of the SCJS (i.e. excludes crimes considered 'out of scope' or unclassifiable).

  A variable called VALID already exists in the VFF data file to identify cases that meet these three criteria. In addition, there is another variable VALIDSJCS which further excludes all sexual assaults and threats identified in the victim forms (as these are not considered reliable for grossing to national level). Users who wish to recreate these variables, or adapt them (for example, to include all cases regardless of timing but exclude those that did not occur in Scotland) should refer to steps 1 to 3 in stage 1, below.

- Stage 2 involves constructing a variable that can be used as a weight to denote how many crimes each case in the VFF represents. This is important for crime victimisation data because some cases reflect only one crime, while others reflect two or more crimes. Two weights are already included in the VFF dataset: NSERIES and NUMINC. This stage indicates how these variables were constructed and how they should be used.
- Stage 3 shows how to conduct the aggregation process. This involves creating the appropriate crime variable, constructing vector loops and aggregating one variable to many variables.

Stage 1: Excluding invalid SCJS crime incidents

1. Exclude any incidents that happened outside Scotland

   Create a new variable ELIGWHERE using the variables QADDALL (a recode of QADD), QLOC, QSCO and QWHERE. ELIGWHERE is given a value of 1 if the incident occurred in the local area and the respondent was living at their current address at the start of the reference period (QLOC=1 and QADDALL=1), or it occurred in their local area but they were living elsewhere in Scotland at the time (QWHERE =1, 2 OR 3), or if it happened away from their local area elsewhere in Scotland (QSCO=1); otherwise it has a value of 0.

```
compute QADDALL=QADD.
recode QADD (missing=1) into QADDALL.

compute ELIGWHERE =0.
if (QLOC=1 and QADDALL=1) ELIGWHERE =1.
if (QSCO=1) ELIGWHERE =1.
if (QWHERE=1 or QWHERE=2 or QWHERE=3) ELIGWHERE =1.
val lab ELIGWHERE 1 'eligible for analysis' 0 'not eligible'.
var label ELIGWHERE 'All variables eligible for analysis in
victimisation rates: occurred in Scotland'.
```

   Out of a possible 5,772 incidents for analysis, 4,955 are considered eligible for analysis based on the fact that they occurred in Scotland.

2. Excluding incidents that happened outside the reference period

   This has to be done in two parts, because some incidents are single incidents (and therefore just have one date) while other incidents involve a series of similar incidents (in which case some of the dates may fall outwith the reference period). Remember here that, because of the continuous nature of the survey, there is a 23 month total reference period overall (ranging from April 2007 to February 2009); however, only those incidents reported during the 12 months prior to the month of interview are recorded for each respondent.

Part 1: single incidents

There are 4,490 single victim forms. For these forms, a new variable ELIGSING is calculated. If the incident occurred in or after April 2007 and before the end of February 2009 (MTHINC2 between 2 and 24), ELIGSING has a value of 1. If the respondent didn't know what month the incident happened, but they are sure that it occurred within a relevant quarter, (QTRINCID between 2 and 22), ELIGSING also has a value of 1. If the respondent didn't know what quarter, but they are sure it occurred within the reference period (YRINCIB=2), ELIGSING also has a value of 1. The syntax below produces 3,858 eligible victim forms.

```
compute ELIGSING  =0.
if ((MTHINC2 ge 2) and (MTHINC2 le 24)) ELIGSING  =1.
if ((QTRINCID ge 2) and (QTRINCID le 22)) ELIGSING  =1.
if YRINCIB=2 ELIGSING  =1.
var lab ELIGSING  'Single incidents: those that occurred in
reference year'.
val lab ELIGSING  1 'yes' 0 'no'.
```

Part 2: series incidents

Checking the valid dates of series incidents involves slightly more steps. There are 1,282 series victim forms. A variable ELIGSER1 is created using the 'DATESER_' variables (which records the month in which series incidents were reported as occurring). If the incident occurred within any of the relevant months from 1st April 2007 to 28th February 2009 (DATESER_02 to DATESER_22 equals 1), ELIGSER1 has a value of 1. Using the syntax below, 1,107 series forms are identified in which all of the incidents occurred in the reference period.

```
compute ELIGSER1 =0.
if (DATESER_02 = 1) or (DATESER_03 = 1) or (DATESER_04 = 1) or
   (DATESER_05 = 1) or (DATESER_06=1)  or (DATESER_07=1)  or
   (DATESER_08=1)  or (DATESER_09=1)  or (DATESER_10=1)  or
   (DATESER_11=1)  or (DATESER_12=1)  or (DATESER_13=1)  or
   (DATESER_14=1)  or (DATESER_15=1)  or (DATESER_16=1)  or
   (DATESER_17=1)  or (DATESER_18=1)  or (DATESER_19=1)  or
   (DATESER_20=1)  or (DATESER_21=1)  or (DATESER_22=1)
   ELIGSER1 =1.
var lab ELIGSER1 'Series incidents 1: those that occurred in
reference year'.
val lab ELIGSER1 1 'yes' 0 'no'.
```

As with single incidents, some respondents may be unclear about the exact date of incidents, but be sure it happened within the correct quarter. Again, syntax can be used to identify these cases using the variables QUART_APR07_JUN07 to QUART_JAN09_MAR09 (see syntax to create

ELIGSER2 below); although, in the 2008/09 VFF no further cases were found.

```
compute ELIGSER2 =0.
if (QUART_APR07_JUN07 ge 1) or (QUART_MAY07_JUL07 ge 1) or
  (QUART_JUN07_AUG07 ge 1) or (QUART_JUL07_SEP07 ge 1) or
  (QUART_AUG07_OCT07 ge 1) or (QUART_SEP07_NOV07 ge 1) or
  (QUART_OCT07_DEC07 ge 1) or (QUART_NOV07_JAN08 ge 1) or
  (QUART_DEC07_FEB08 ge 1) or (QUART_JAN08_MAR08 ge 1) or
  (QUART_FEB08_APR08 ge 1) or (QUART_MAR08_MAY08 ge 1) or
  (QUART_APR08_JUN08 ge 1) or (QUART_MAY08_JUL08 ge 1) or
  (QUART_JUN08_AUG08 ge 1) or (QUART_JUL08_SEP08 ge 1) or
  (QUART_AUG08_OCT08 ge 1) or (QUART_SEP08_NOV08 ge 1) or
  (QUART_OCT08_DEC08 ge 1) or (QUART_NOV08_JAN09 ge 1) or
  (QUART_DEC08_FEB09 ge 1) or (QUART_JAN09_MAR09 ge 1)
  ELIGSER2 =1.
var lab ELIGSER2 'Series incidents 2: those that occurred in
reference
        year'.
val lab ELIGSER2 1 'yes' 0 'no'.
```

A final check variable, ELIGSER3, is created to ensure whether – if the respondent is unsure of the dates of all the incidents – they are clear that the most recent incident occurred within a month (MTHRECIN) or quarter (QTRRECIN) or simply the year (YRINC) of the reference period. If not, the case is considered invalid and excluded from analysis. This final check reduces the number of eligible series victim forms from 1,107 to 1,103.

```
compute ELIGSER3 = ELIGSER2.
if (MTHRECIN =1) ELIGSER3 =0.
if (QTRRECIN=1) ELIGSER3 =0.
if (YRINC=1 or YRINC=4 or YRINC=5) ELIGSER3 =0.
var lab ELIGSER3 'Series incidents 3: those that occurred in
reference year'.
val lab ELIGSER3 1 'yes' 0 'no'.
```

Finally, it is necessary to create a count of the number of incidents occurring within the reference period. The count variables for each quarter (QUART_APR07_JUN07 to QUART_JAN09_MAR09) are recoded so that any value not between 1 and 96 is recoded to 0, which means very high values are included here).

```
recode QUART_APR07_JUN07 QUART_MAY07_JUL07
  QUART_JUN07_AUG07 QUART_JUL07_SEP07
  QUART_AUG07_OCT07 QUART_SEP07_NOV07
  QUART_OCT07_DEC07 QUART_NOV07_JAN08
  QUART_DEC07_FEB08 QUART_JAN08_MAR08
  QUART_FEB08_APR08 QUART_MAR08_MAY08
```

```
QUART_APR08_JUN08 QUART_MAY08_JUL08
QUART_JUN08_AUG08 QUART_JUL08_SEP08
QUART_AUG08_OCT08  QUART_SEP08_NOV08
QUART_OCT08_DEC08 QUART_NOV08_JAN09
QUART_DEC08_FEB09 QUART_JAN09_MAR09 (1 thru 96=copy)
(else=0) into SEQT_1 SEQT_2   SEQT_3 SEQT_4 SEQT_5 SEQT_6
SEQT_7 SEQT_8 SEQT_9 SEQT_10 SEQT_11 SEQT_12 SEQT_13
SEQT_14 SEQT_15 SEQT_16 SEQT_17 SEQT_18 SEQT_19
SEQT_20 SEQT_21 SEQT_22.
```

The next step is to select only those cases that are eligible, according to ELIGSER3, and add the values together. This creates a count variable ELIGSER3N which ranges from 0 (i.e. the cases that are not eligible) to 96 (i.e. the highest number of incidents recorded in any series).

```
do if ELIGSER3=1.
compute ELIGSER3N =SEQT_1 + SEQT_2 + SEQT_3 + SEQT_4 +
SEQT_5 + SEQT_6 + SEQT_7   + SEQT_8 + SEQT_9 + SEQT_10 +
SEQT_11 + SEQT_12 + SEQT_13 + SEQT_14 + SEQT_15  + SEQT_16 +
SEQT_17 + SEQT_18 + SEQT_19 + SEQT_20 + SEQT_21 + SEQT_22.
end if.
```

If the respondent is very clear that cases happened within the reference period (so the case is eligible according to ELIGSER3), but they either could not or would not specify when they occurred (so ELIGSER3N has a value of 0), then these cases are given an "imputed" value of 2. This was done by the survey company on the basis of discussions with Scottish Government and following the procedure used for the British Crime Survey.

```
do if ELIGSER3=1 and eligser3n=0.
recode ELIGSER3N (0=2).
end if.
var lab ELIGSER3N 'Series incidents 3: how many in reference
year'.
```

3. Excluding crime types that are not considered valid for SCJS

During the coding process, there are some crimes that cannot be classified into any of the standard SCJS crime codes (see the variable OFFENCE for details of what these are). In some cases, there may not be sufficient information on which to make a crime classification; for example, damage to a vehicle may have been accidental rather than malicious; or property may have been lost rather than stolen. In other cases, crimes may be considered outwith the scope of the survey.

There are two types of crime for which information is collected, but these are not used to calculate national estimates. The first is sexual offences. The number of sexual offences reported at the victim screener stage is

extremely small, and it is known that this is not a reliable method of collecting accurate data about such incidents; therefore, national estimates would be extremely misleading. The second is threats, which are not used to calculate national estimates due to the problems of establishing whether a crime actually occurred or not.

Variables can be created to exclude those crime types that are not considered valid for SCJS. To create a variable to exclude those crime types that are not considered classifiable or within scope of the SCJS, the following syntax could be used to create a new variable 'ELIGTYPE':

```
compute ELIGTYPE = 1.
if any(OFFENCE,1,6,16,22,23,28,33,40,42,43,47,52,53,54,59,60,
61,62,63) ELIGTYPE = 0.
var lab ELIGTYPE 'Incidents classifiable and within scope of SCJS'.
val lab ELIGTYPE 1 'Eligible' 0 'Not eligible'.
```

To also exclude sexual assaults and threats from analysis of the SCJS data, the syntax shown above can be amended to include the following values of OFFENCE following the 'if any' statement: 8,9,10,11,12,13,14,15 (for sexual assaults) and 55,56,57,58 (for threats).

4. Creating an overall eligibility variable.

As noted above, there already exist two eligibility variables in the VFF that allow users to exclude invalid cases. VALID has a value of 0 for cases that occurred outwith Scotland, outside the 12 month reference period and which are out of scope or unclassifiable. VALIDSCJS further excludes all sexual assaults and threats.

By combining in different ways the three sets of syntax noted above, it is possible for users to create other eligibility variables according to their needs. So, for example, an eligibility variable (ELIGALL) that simply identifies whether incidents occurred in Scotland (ELIGWHERE=1) and during the relevant time period (ELIGSING=1 AND ELIGSER3=1), but which does not exclude cases on the basis of their offence coding, could be created using the syntax shown below. This syntax identifies a total of 4,800 eligible victim forms from 5,772 (i.e. 16.8% of forms are excluded from further analysis).

```
compute ELIGALL=0.
if (ELIGWHERE=1) and ((ELIGSING=1) or ((ELIGSER3=1)))
ELIGALL=1.
var lab ELIGALL'Crimes within Scotland and during 12 month
reference period'.
val lab ELIGALL1 'Eligible' 0 'Not eligible'.
```

Stage 2: Creating series weights

5. Uncapped weight

When aggregating crime victimisation data to individual level, it is necessary to use a series weight to reflect how many incidents of victimisation each variable represents. There are two variables already contained in the VFF dataset which can be used for this weighting purpose. The first is NSERIES which denotes the number of incidents represented by each series victim form (n=1282). NSERIES has a value of n for all series incident crimes, where n is the number of incidents occurring in the series. There are 166 cases with a value of zero in NSERIES, and these are cases where the crimes occurred during the month of interview and, therefore, not within the full 12 month reference period (in other words, they are weighted to zero as they are not included in the crime count). NSERIES does not, however, include the single incidents, which may be important for further analysis. If users wish to have a weight that also takes account of single incidents, this can be easily done using the following syntax.

```
if (PINCI=1) NSERIES =1.
```

6. Capped weight

NSERIES and NSERELIG, described above, are uncapped weights, meaning that their highest value is the highest number of crime incidents reported in any one victim form. However, it is common practice in national crime surveys to 'cap' series incidents at a certain level, so as to prevent small numbers of repeat victims over-inflating the national crime estimates (and widening the confidence intervals). Therefore, both the Scottish and British surveys use a series weight which caps series incidents at 5. This means that victims who experienced 6 or more series incidents of victimisation will have a limit of 5 set on them. A variable capping NSERIES at a maximum of 5 is included in the VFF data file. This variable is NUMINC, and can be replicated using the syntax shown below. This process results in 12% of all series victim forms being capped at 5.

```
recode NSERIES (0 thru 5=copy) (5 thru hi=5) into NUMINC.
var lab NUMINC  'Number of incidents in series (capped at 5)'.
```

Again, the NUMINC variable included does not take account of the single incidents, so the following line of syntax would have to be run to represent a weight for all cases:

```
if (PINCI=1) NUMINC =1.
```

Users need to consider whether they wish to use a capped or an uncapped weight before progressing on to stage 3 of the data merging process, discussed below. To replicate SCJS data, the capped weight should be used.

Stage 3: Aggregating one variable to many variables

7. Create a suitable variable for aggregating

Each incident (or series of incidents) reported by a victim to the SCJS is given a specific offence code (see section 2.5 on the victim form file and Table A1 in the Appendix). In the VFF, the variable OFFENCE represents the offence code for each victim form. These are very specific codes and there are many of them, which means that certain codes (for uncommon types of crime) are rarely used. Therefore, they tend to be grouped together into general crime categories, such as vandalism, assault and housebreaking (see Tables A1 and A2 in the Appendix).

Suitable syntax for grouping variables into 15 broad categories covered by the SCJS was provided earlier, in section 2.5 (see syntax example 2). Note that in the SCJS publications, not all crime data collected by the survey are reported as some are considered unreliable for national estimates (e.g. data on sexual assaults and threats) and others are classified as out of scope of the survey or non-classifiable (for further details on this, see Page et al, 2009: 66-68). Depending on the user's needs, crime types can be classified in a range of different ways.

For this example, the variable OFFENCE is recoded into a new variable OFFNEW which contains three broad crime categories that are commonly used in the SCJS reports because they are comparable to police recorded crimes. The first is 'vandalism' which includes fire raising, vandalism to a motor vehicle, vandalism to the home and other vandalism. The second is 'acquisitive crime' which includes theft of a motor vehicle, bicycle theft and attempted or actual housebreaking. The third is 'violence' which includes all minor assault, serious assault and violent theft (robbery). All other crime codes are set to missing for the purposes of this example.

```
Recode OFFENCE (49,48,50,51=1) (25,26,27,34,36,38=2)
  (2,3,4,5,7,17,18=3) (else=sysmis) into OFFNEW.
var lab OFFNEW 'Three new types of victimisation'.
val lab OFFNEW 1 'Vandalism' 2  'Acquisitive crime' 3 'Violence'.
```

A frequency analysis should provide the output shown in Figure 3.4 for this new variable (this is based on unweighted data). As can be seen, 3220 cases in the dataset have been excluded because they do not fall into one of these three broad crime categories. Of the remaining 2552 cases, 61.2% are vandalism incidents, 25.4% are violence incidents and acquisitive crime incidents account for 13.3% of the total.

Figure 3.4: SPSS output of frequency of new victimisation types

OFFNEW Three new types of victimisation

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00 Vandalism	1563	27.1	61.2	61.2
	2.00 Acquisitive crime	340	5.9	13.3	74.6
	3.00 Violence	649	11.2	25.4	100.0
	Total	2552	44.2	100.0	
Missing	System	3220	55.8		
Total		5772	100.0		

8. Create a new dataset for merging to the RF

   Before progressing further with the aggregation process, it is useful to create a new dataset containing only those variables (and cases) that you wish to aggregate. In the example syntax shown below, a 'select if' command is used to exclude the cases taking place outside Scotland and outwith the reference period (we do not need to worry about the non-classifiable cases here as we are only using variables that do have a definite code). Users who are not concerned with excluding these cases should ignore this command.

   Create a new data file using the 'save outfile' command, and using the 'keep' sub-command to save only a small number of variables. In this case, all we need to save to the new data file are four variables:

   - the respondent's unique serial number (SERIAL);
   - the unique serial number of each victim form (VSERIAL);
   - the crime type variable (OFFNEW);
   - and the weighting variable which denotes how many incidents each victim form represents. Here you could use either NSERIES (uncapped weight) or NUMINC (capped at 5).

   It is very important to remember that neither of the weighting variables (NSERIES OR NUMINC) contained in the VFF take account of the single incidents; therefore, this needs to be done before the following steps are followed (otherwise none of the single incidents will be counted). In this example, the capped weight NUMINC, with the single incidents given a value of one (see step 6), is used as we wish to replicate variables already included in the RF dataset.

   You must save the work you have done to the VFF dataset before running the syntax shown below; otherwise, as soon as you run the 'select if' command, you will lose a large proportion of your cases and this will be irretrievable.

```
select if (ELIGALL=1) .
save outfile="insert your file pathway here"
   / keep SERIAL VSERIAL OFFNEW NUMINC.
```

Open your new data file using the 'get file' command shown below, and remember to close the original victim file (without saving it if you have used the 'select if' command, as this will permanently delete all cases where VALID did not equal 1). You should have 4,800 cases in the new data file.

```
get file="insert your file pathway here".
```

9. Aggregate the OFFNEW variable to individual level

To merge the victimisation data on to the main data file, the data need to be aggregated from multiple records per victim to one record per respondent – taking account of the fact that each victim may have up to five victim forms AND taking account of the fact that each victim form may represent multiple incidents. This can be done using vector loops.

To start the process of aggregation, sort the cases on the variable you want to use for aggregation (in this case SERIAL) and then use the 'casestovars' command to group the variables. Because there are a maximum of 5 victim forms per person (i.e. per SERIAL) five new variables will be created for each of the variables in the ungrouped dataset. You can check this by looking at the dataset.

```
sort cases by SERIAL.
casestovars
  /id = SERIAL
  /groupby = variable.
```

After running this command there should be 3,270 records in the dataset - each of which corresponds to an individual. Individuals who had multiple records will now have multiple new variables instead.

10. Vector loops to create separate variables for each crime type

The next stage is to create a separate crime classification variable for each of the 3 values of OFFNEW. This can be done using a vector command (which takes one variable and splits it into several). A loop command is also used to ensure that the data from all five victim forms are taken into account in creating the new variables. Vector loops are just a shorthand way of getting the same command run multiple times – in this case five times.

The three new variables require a variable label preface. In this case, the preface 'NOFF' is used to stand for 'number of offences'. Since the

variable OFFNEW has only 3 values, three new variables will be created – NOFF1 (for vandalism), NOFF2 (for acquisitive crime) and NOFF3 (for violence). The labels are really irrelevant – you could call them anything – and each variable can be renamed later (see stage 12).

The value of each new variable is the combined value of all the victim forms for that crime type (bearing in mind that each form is capped at a maximum of 5). So if OFFNEW=1 (vandalism), NOFF1 is equal to the total number of vandalism incidents reported by each individual, which could range from 1 (i.e. one single incident) to 25 (i.e. 5 series incidents capped at 5).

```
vector NOFF(3) .
recode NOFF1 to NOFF3 (missing = 0) .
vector OFFNEW = OFFNEW.1 to OFFNEW.5 .
vector series = NUMINC.1 to NUMINC.5 .
loop #i = 1 to 5 .
  if OFFNEW (#i) = 1 NOFF1 = NOFF1 + series(#i) .
  if OFFNEW (#i) = 2 NOFF2 = NOFF2 + series(#i) .
  if OFFNEW (#i) = 3 NOFF3 = NOFF3 + series(#i) .
end loop .
EXE.
```

Observing the dataset, you will see that you now have three new variables: NOFF1, NOFF2 and NOFF3. Users wishing to run the same analysis using uncapped data should replace the variable NUMINC with NSERIES at step 8.

11. Merging the NOFF variables onto the RF dataset

The three NOFF variables can now be merged onto the RF dataset. This can be done using exactly the same procedures as shown in section 3.4.1, using the RF dataset as the 'active dataset' and merging the new variables on to that.

12. Recoding the NOFF variables

Once the NOFF variables have been merged onto the RF dataset, they need to be recoded because they only include valid values for the victims. Any non-victims will all be coded as 'system missing', so these should be recoded as 0. Then variable labels can be added, as shown below.

```
recode NOFF1 NOFF2 NOFF3 (1 thru hi = copy) (else=0) .
var lab noff1 'Number of vandalism incidents'  /
  noff2 'Number of acquisitive crime incidents'  /
  noff3 'Number of violence incidents'.
```

The NOFF variables are count variables, but you will also want a prevalence measure (i.e. whether each respondent is a victim of crime or

not). This can easily be achieved by recoding the count variables. Here we recode the variables using the prefix POFF to show that it is a prevalence variable (but again, the labels could be anything). This time value labels should also be added: 1 is coded 'yes' and 2 'no'.

```
recode NOFF1 NOFF2 NOFF3 (1 thru hi = 1) (0=2) into POFF1
POFF2 POFF3.
var lab POFF1 'Any vandalism incidents'  /
  POFF2 'Any acquisitive crime incidents'  /
  POFF3 'Any violence incidents'.
val lab POFF1 to POFF3 1 "Yes" 2 "No" .
```

13. Checking the new variables against the derived variables in the RF dataset

The final step in this process is to check the new variables to ensure they look correct. In this example, we have recreated three of the incidence variables that are already contained in the RF dataset so we can check to ensure that no errors have occurred. This can be done by running a simple frequency check in the RF data file on the three NOFF variables against their equivalent incidence variables (INCVAND, INCACQUIS and INCVIOLENT) and the three POFF variables against their equivalent prevalence variables (PREVVAND, PREVACQUIS and PREVVIOLENT) . By running the following syntax, you should be able to replicate the results shown table 3.5 below.

```
weight off.
fre POFF1 POFF2 POFF3.

missing values NOFF1 NOFF2 NOFF3 (0).
fre NOFF1 NOFF2 NOFF3.
```

Figure 3.5: Summary of analysis showing prevalence and incidence of vandalism, acquisitive crime and violence

	Vandalism	Acquisitive Crime	Violence
Number of respondents who were victims	1326	308	529
% of respondents who were victims (prevalence)	8.3	1.9	3.3
% of victims who were victimised 1 or more times (incidence):			
1 time	65.9	85.7	63.1
2 times	18.8	8.4	18.5
3 times	7.0	3.6	7.0
4 times	4.1	1.3	4.2
5+ times	4.1	0.1	7.2

Note: The data shown in this table are not weighted as they are for illustrative purposes only. The household weight should be applied for vandalism and acquisitive crimes, and the individual weight should be applied for violence.

3.4.3 Merging variables from the SCF to the RF

Merging data from the self-completion data file (SCF) to the respondent file (RF) is more straightforward than the examples above because both files have the individual level as the unit of analysis. It is important to remember, however, that not all of those who responded to the main survey also completed a self-completion questionnaire. Therefore, although the response rate to the self-completion questionnaire was 69% of survey respondents, this represents only 48% of the original sample as a whole. For this reason, a separate set of weights are calculated for the self-completion questions; and if data are merged from the SCF to the RF, it is important to also copy the self-completion weights to the RF.

Approach/analysis

To illustrate this data merging process, we will use an example that requires data from both the self-completion questionnaire and the main respondent survey. The question is: do respondents from the Lesbian, Gay, Bisexual and Transexual (LGBT) community feel less safe walking alone in their local area after dark than other members of the Scottish public? To answer this question, we need the variable QSFDARK (how safe do you feel walking alone in your local area after dark?) from the respondent data file; and the variable SEXORIEN (sexual orientation) from the self-completion data file. It is important also to transfer the scaled individual weight from the self-completion file in order that the analysis can be appropriately weighted.

Syntax/output

Start by opening the RF data file, and run a weighted frequency analysis of the variable QSFDARK to produce the output shown below. There are a small percentage of people who said 'don't know' so these can be set to missing (this is not essential - you could retain them in the analysis if you were interested in this group of people).

```
weight by WGTGINDIV_SCALE.
fre QSFDARK.
```

qsfdark QSFDARK: How safe respondent feels walking alone in local area after dark

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	-1.00 Don't know	220	1.4	1.4	1.4
	1.00 Very safe	4680	29.2	29.2	30.6
	2.00 Fairly safe	5917	37.0	37.0	67.6
	3.00 A bit unsafe	3295	20.6	20.6	88.2
	4.00 Very unsafe	1889	11.8	11.8	100.0
	Total	16000	100.0	100.0	

Open the self-completion data file, and run a weighted frequency of the variable SEXORIEN using the scaled individual weight. The output below shows that few people reported being anything other than heterosexual; but, a substantial minority of respondents people did not wish to answer the question.

```
weight by WGTGINDIV_SC_SCALE.

fre SEXORIEN.
```

sexorien SEXORIEN: Sexual orientation

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	-2.00 Don't wish to answer	118	1.1	1.1	1.1
	1.00 Heterosexual (straight)	10616	96.7	96.7	97.8
	2.00 Gay or lesbian	90	.8	.8	98.6
	3.00 Bisexual	63	.6	.6	99.2
	4.00 Other	87	.8	.8	100.0
	Total	10974	100.0	100.0	

There is not much that we can determine from those who did not answer, so we will recode this variable so that the gay, lesbian, bisexual and 'other' responses are combined into one category, and the 'don't wish to answer' group is set to missing. This requires a recode of the SEXORIEN variable into SEXORIEN2. The syntax below would achieve this aim:

```
recode SEXORIEN (1=copy) (2,3,4=2) (else=sysmis) into SEXORIEN2.
var lab SEXORIEN2 'Sexual orientation - recoded'.
val lab SEXORIEN2 1 'heterosexual (straight)' 2 'LGBT'.

fre SEXORIEN2.
```

The output below shows that the LGBT respondents represent a very small percentage of the sample, so we have to be cautious about drawing conclusions about the LGBT community as a whole.

SEXORIEN2 Sexual orientation - recoded

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00 heterosexual (straight)	10616	96.7	97.8	97.8
	2.00 LGBT	240	2.2	2.2	100.0
	Total	10856	98.9	100.0	
Missing	System	118	1.1		
Total		10974	100.0		

To merge the data on sexual orientation into the respondent file, we will use a procedure similar to that used in example 3.4.1, above. For this example, the command windows method is demonstrated; however, it would also be feasible to use the match files syntax method. The steps are shown below.

1. Open both the SCF and the RF data files in SPSS.

2. Sort both files ascending on the variable SERIAL using the following syntax:

   Sort cases by SERIAL(A).

3. Ensure that the RF is the 'active dataset' (i.e. the one you have open in front of you).

4. In the main menu, go to 'Data', click on 'Merge files', and then 'Add variables'.

5. Click on 'An active dataset' and select the SCF from the list. Press continue.

6. A new window, 'Add variables from…' will open (see Figure 2.2). The dialogue box 'new active dataset' contains the names of all the variables in the primary dataset (RF – marked *) and all the additional variables from the secondary dataset (SCF – marked +). You need to transfer all those variables from the SCF that you do not wish to merge on to the RF back into the 'excluded variables' list. The two variables you wish to merge onto the RF (SEXORIEN2 and WGTGINDIV_SC_SCALE) are both near the bottom of the list, so you can just select all the others marked + and transfer them back to the excluded variables. (Use the shift key to select multiple variables rather than doing them one by one)

7. Once SEXORIEN2 and WGTGINDIV_SC_SCALE are the only variables from the SCF that are included in the 'new active dataset' dialogue box, select 'match cases on key variable in sorted files'.

8. In the 'Excluded variables' dialogue box find SERIAL(+) from the excluded variables list and press the lower, right-pointing arrow to place this variable in the 'key variables' dialogue box. This allows you to match the files on the basis of the individual's unique identifier.

9. Click on 'non-active dataset is keyed table' which tells SPSS to match the cases in the SCF to those in the RF. (Where there are missing data for individual cases, this will be set to system missing). Press OK.

10. As a matter of default when you run this analysis a message will appear stating 'Warning: Key match will fail if data are not sort in ascending order of key variables'. (This is to remind you that both files need to be sorted in the same order by the same variable - SERIAL).

11. Press OK again and SPSS will merge the two variables from the SCF onto the RF.

Before progressing further you should run a frequency analysis of the variable SEXORIEN2 in the respondent file, using the self-completion individual weight, to check that the variables have transferred successfully (not shown here). The next step is to cross-tabulate the variable QSFDARK with sexual orientation (again, using the self-completion individual weight) to determine if the LGBT respondents were more concerned about walking in their local area after dark.

```
weight by WGTGINDIV_SC_SCALE.

crosstabs
  /tables=QSFDARK by SEXORIEN2
  /format=avalue tables
  /statistics=chisq gamma
  /cells=count column
  /count round cell.
```

The table below indicates that there was some difference between the heterosexual and the LGBT respondents to the question about safety after dark. The LGBT respondents were less likely to say that they felt very safe, and more likely to say they felt fairly safe, compared to the heterosexual respondents. However, the heterosexual respondents were a little more likely to say they felt very unsafe. Statistical tests (not shown) suggest the difference is significant, although perhaps not so extreme as to require specific intervention.

qsfdark QSFDARK: How safe respondent feels walking alone in local area after dark * sexorien2 Sexual orientated - recoded Crosstabulation

			sexorien2 Sexual orientated - recoded		
			1.00 heterosexual (straight)	2.00 LGBT	Total
QSFDARK: How safe respondent feels walking alone in local area after dark	1.00 Very safe	Count	3145	61	3206
		% within sexorien2	29.9%	25.5%	29.8%
	2.00 Fairly safe	Count	3956	110	4066
		% within sexorien2	37.6%	46.0%	37.8%
	3.00 A bit unsafe	Count	2150	49	2199
		% within sexorien2	20.4%	20.5%	20.4%
	4.00 Very unsafe	Count	1277	19	1296
		% within sexorien2	12.1%	7.9%	12.0%
Total		Count	10528	239	10767
		% within sexorien2	100.0%	100.0%	100.0%

Section 4: Resources for survey users

The following links are intended to provide some quick and easy pointers to resources that may be of benefit to users of the SCJS 2008/09, or other large-scale surveys more generally.

Further information about the Scottish Crime and Justice Survey can be obtained from the Scottish Government website:
http://www.scotland.gov.uk/Topics/Statistics/Browse/Crime-Justice/crime-and-justice-survey

The Research Development and Statistics department of the Home Office provides some useful links and resources relating to the British Crime Survey and other crime related surveys:
http://rds.homeoffice.gov.uk/rds/bcs1.html

The Scottish Centre for Crime and Justice Research website provides useful information and links on crime and criminal justice related issues.
http://www.sccjr.ac.uk/

The Applied Quantitative Methods Network provides a range of resources that may be of use to analysts within Scotland, including a consultation service:
http://www.aqmen.ac.uk/

Survey Resources Network is a new service funded by the Economic and Social Research Council (ESRC) to coordinate and develop ESRC activities related to survey research methods.
http://surveynet.ac.uk/

ESDS learning and teaching website; uses datasets to demonstrate. Includes a section on 'Use of Data', 'Resources', and a 'Step-by-Step' Guide.
http://www.esds.ac.uk/resources/E5.asp

ESDS also produces a number of guides which are available online, that both introduce and provide guidance on various ESDS services.
http://www.esds.ac.uk/support/onlineguides.asp

Guides to statistical analysis of ESDS governmental datasets can be found here:
http://www.esds.ac.uk/government/resources/analysis/

Appendix

Table A1: Offence codes and crime groups in the SCJS 2008/09 (using variable OFFENCE in the victim form file)

Original variable: OFFENCE		Derived crime groups
Values	**Value labels**	
2	11 Serious assault	Assault
3	12 Minor assault	
4	14 Serious assault and fire raising	
5	15 Serious assault and housebreaking	
7	21 Attempted assault	
17	41 Robbery	Robbery
18	42 Attempted robbery	
19	43 Snatch theft from the person	Other personal theft (excluding robbery)
20	44 Other theft from the person	
21	45 Attempted theft from the person	
41	67 Other theft	
46	73 Other attempted theft	
25	51 Housebreaking in a dwelling (nothing taken)	Housebreaking
26	52 Housebreaking in a dwelling (something taken)	
27	53 Attempted housebreaking in a dwelling	
24	50 Attempted housebreaking to non-connected domestic garage / outhouse	Other attempted theft (including bicycle theft)
29	55 Theft in a dwelling	
30	56 Theft from a meter	
31	57 Housebreaking from non-connected domestic garage/outhouse – nothing taken	
32	58 Housebreaking from non-connected domestic garage/outhouse – something taken	
38	64 Theft of pedal cycle	
39	65 Theft from outside dwelling (excluding theft of milk bottles)	
34	60 Theft of car/van	All motor vehicle theft
35	61 Theft from car/van	
36	62 Theft of motorbike, motor scooter or moped	
37	63 Theft from motorbike, motor scooter or moped	
44	71 Attempted theft of/from car/van	
45	72 Attempted theft of/from motorcycle, motor scooter or moped	
48	80 Fire raising	Vandalism
49	82 Vandalism to a motor vehicle	
50	84 Vandalism to the home	
51	86 Other vandalism	

Source: Adapted from Page et al (2009), p68

Table A2: Offence codes and aggregate crime groups

Offence codes	Aggregate crime group
60 Theft of car / van 64 Theft of pedal cycle 51 Housebreaking in a dwelling (nothing taken) 52 Housebreaking in a dwelling (something taken) 53 Attempted housebreaking in a dwelling	Acquisitive crime
11 Serious assault 12 Minor assault 14 Serious assault and fire raising 15 Serious assault and housebreaking 21 Attempted assault 41 Robbery 42 Attempted robbery	Violent crime
60 Theft of car / van 64 Theft of pedal cycle 51 Housebreaking in a dwelling (nothing taken) 52 Housebreaking in a dwelling (something taken) 53 Attempted housebreaking in a dwelling 11 Serious assault 12 Minor assault 14 Serious assault and fire raising 15 Serious assault and housebreaking 21 Attempted assault 41 Robbery 42 Attempted robbery 80 Fire raising 82 Vandalism to a motor vehicle 84 Vandalism to the home 86 Other vandalism	All comparable crimes (i.e. comparable with police recorded crime statistics)
All crime codes in T able A1	All SCJS survey crimes

Note: Crime codes for sexual assault and threats are also included in the SCJS 2008/09; however, these are not included in any of the crime groups or aggregate crime groups.

References

Anderson, S. (1999) 'Crime Statistics and the 'Problem of Crime' in Scotland', in Hutton, N. and Duff, P. (Eds) *Criminal Justice in Scotland.* Dartmouth: Ashgate.

Anderson, S. and Leitch, S. (1996) *Main Findings from the 1993 Scottish Crime Survey.* Edinburgh: Scottish Office Central Research Unit.

Brown, M. and Bolling, K. (2007) *2006 Scottish Crime and Victimisation Survey: Main Findings*. Edinburgh: Scottish Government Social Research.

Chambers, G. and Tombs, J. (1984) *The British Crime Survey Scotland*. Edinburgh: Scottish Office Central Research Unit.

Hope, S. (2005) *Scottish Crime and Victimisation Survey Calibration Exercise: A comparison of survey methodologies.* Edinburgh: Scottish Executive.

Hope, S. (2007) *Scottish Crime and Victimisation Survey Calibration 2004: Research Findings.* Edinburgh: Scottish Government Social Research Findings No.86.

Lynn, P. (1996) 'Who responds to the British Crime Survey?' *Paper presented to the Fourth International ISA Conference on Social Science Methodology*.

Lynn, P. and Elliot, D. (2000) *The British Crime Survey: A Review of Methodology.* London: National Centre for Social Research.

McAra, L. (2008) Crime, Criminology and Criminal Justice in Scotland. *European Journal of Criminology,* 5: 481

McCaig, E. and Leven, T. (2003) *Fundamental review of the Scottish Crime Survey*, Edinburgh: Scottish Executive Social Research.

MacLeod, P., Page, L., Kinver, A., Iliasov, A., Littlewood, M. and Williams, R (2009) *2008/09 Scottish Crime and Justice Survey: First Findings* Edinburgh: Scottish Government

McVie, S. and Norris, P. (2007) Review of Scottish Crime and Justice Survey Questionnaire. SCCJR briefing report produced for the Scottish Executive.

McVie, S., Campbell, S. and Lebov, K. (2004) *Scottish Crime Survey 2003.* Edinburgh: Scottish Executive Social Research.

MVA (1998) *Main Findings from the 1996 Scottish Crime Survey.* Edinburgh: Scottish Office Central Research Unit.

MVA (2002) *The 2000 Scottish Crime Survey: Overview Report.* Edinburgh: Scottish Executive Social Research.

Nicholas, S., Kershaw, C. and Walker, A. (2007) *Crime in England and Wales 2006/07*. Home Office Statistical Bulletin. London: Home Office.

Norris, P. and McVie, S. (2009) A Review of the 2006 Scottish Crime and Victimisation Survey: Victim File and Victimisation Rates. SCCJR briefing report produced for the Scottish Government.

Page, L., MacLeod, P., Rogers, J., Kinver, A., Iliasov, A. and Tseloni, A (2009) *2008/09 Scottish Crime and Justice Survey: Technical Report* Edinburgh: Scottish Government

Page, L., MacLeod, P., Kinver, A., Iliasov, A. and Yoon, P. (2010a) *2009/10 Scottish Crime and Justice Survey: Main Findings* Edinburgh: Scottish Government

Page, L., MacLeod, P. and Twist, N. (2010b) *2009/10 Scottish Crime and Justice Survey: Technical Report* Edinburgh: Scottish Government

Payne, D. (1992) *Crime in Scotland: findings from the 1988 British Crime Survey.* Edinburgh: Scottish Office Central Research Unit.