



Crime Survey
for England & Wales

Crime Survey for England and Wales

Technical Report 2019/20

Volume One

KANTAR



Contents

1.	Background	3
2.	Sample Design	6
3.	Questionnaire content and development	154
4.	Developing a risk rating system	310
5.	Fieldwork	332
6.	Response rate and reasons for non-response: core sample	41
7.	Data Processing	53
8.	Data Output	542
9.	Weighting	688
10.	Comparing key survey variables with the population	683

1. Background

1.1 Introduction to the Crime Survey for England and Wales

The Crime Survey for England and Wales (CSEW) is a well-established study and one of the largest social research surveys conducted in England and Wales. The survey was first conducted in 1982 and ran at roughly two-yearly intervals until 2001, when it became a continuous survey¹. Prior to April 2012, the survey was known as the British Crime Survey (BCS) and conducted on behalf of the Home Office. From April 2012 responsibility for the survey transferred to the Office for National Statistics (ONS) and it became known as the Crime Survey for England and Wales (CSEW). Since 2001, Kantar's Public Division has been the sole contractor for the survey.

The CSEW is primarily a survey of **victimisation** in which respondents are asked about their experiences of both **household crimes** (e.g. burglary, vehicle crime) and **personal crimes** (e.g. robbery, snatch theft). Household crimes may have happened to anyone in the household, while personal crimes are only counted if they relate to the individual being interviewed. The reference period for all interviews relates to incidents that have happened in the last 12 months before the date of interview. Although there have been changes to the design of the survey over time, the wording of the screener questions that are asked to elicit respondents' experiences of victimisation have been consistent over the lifetime of the survey. In 2015-16 an additional set of screener questions was added to measure fraud and cybercrime.

Respondents are asked about their experience of crime, irrespective of whether they reported these incidents to the police. As such, the CSEW provides a record of peoples' experiences of crime which is unaffected by variations in reporting behaviour of victims or variations in police practices of recording crime. The CSEW and police recorded figures are two complementary series, which together provide a better picture of crime than can be obtained from either series alone.

Since the survey became continuous in 2001 there have been a few significant changes to the design of the survey. Where changes have been incorporated these have been described in detail in the relevant technical reports. The most significant changes include:

- Between 2004-05 and 2011-12, the core sample size was increased from 37,000 to 46,000, with a target of at least 1,000 interviews in each Police Force Area (PFA).
- Long-standing boost samples of Black and Asian respondents (3,000 sample boost per year) and 16 to 24 year olds (2,000 sample boost per year) were dropped in 2006-07 and 2008-09 respectively.
- In 2009-10, after an extensive development period, the survey was extended to cover young people aged 10 to 15 with a target sample size of 4,000 per year (reduced to 3,000 from 2012-13 onwards)². The first results for this age group were published in June 2010³ as experimental statistics and estimates of victimisation among children are now presented alongside the adult crime statistics.

¹ Previous sweeps of the British Crime Surveys were carried out in 1982, 1984, 1988, 1992, 1994, 1996, 1998 and 2000.

² A feasibility study was carried out before the survey was extended to this age group. See Pickering, K., Smith, P., Bryson, C. and Farmer, C. (2008) British Crime Survey: options for extending the coverage to children and people living in communal establishments. Home Office Research Report 06. London: Home Office.

³ Millard, B. and Flatley, J. (2010) Experimental statistics on victimisation of children aged 10 to 15: Findings from the British Crime Survey for the year ending December 2009. Home Office Statistical Bulletin 11/10.

- In 2012-13 the core adult sample size was reduced from 46,000 to 35,000. In the same year a new sampling approach was adopted based around a three-year unclustered sample design.
- In 2015-16 the questionnaire was updated to include measures of fraud and cybercrime following an extensive development phase, including a large-scale field test. A methodological note about the development of the fraud measures and the field trial was published in 2015 and the questions were put on the survey from October 2015⁴.

Since 2012-13, the core adult (16+) sample size has been approximately 35,000 interviews, conducted across the year. From the 2018-19 survey the target sample size was reduced slightly to 34,500 interviews per year. The current survey is designed to achieve a minimum of around 625 adult interviews in each PFA in England and Wales (reduced from 650 in previous years). The survey is also designed to interview a nationally representative sample of around 3,000 children aged 10 to 15.

Many of these changes to the survey have been driven by various reviews looking at how crime statistics can be improved. These reviews have generally covered both the statistics produced from the survey and police recorded crime. They include:

- National Statistician's Review of Crime Statistics: England and Wales, June 2011
- UK Statistics Authority Assessment of Crime Statistics, January 2014
- Public Administration Select Committee inquiry, April 2014
- Inspection of Crime Data Integrity by Her Majesty's Inspectorate of Constabulary, October 2014

ONS regularly publish updates on the progress that has been made in implementing some of the recommendations from these reviews⁵.

1.2 Outputs from the CSEW

Following the move of the processing and publication of crime statistics to ONS from the Home Office, the standard quarterly releases were extended to include more long-term trends and other data sources.

In addition to the regular quarterly publication, ONS publish additional thematic publications and articles on particular aspects of crime. Recent examples of thematic reports and articles based on CSEW data include:

- Domestic abuse in England and Wales: November 2019
- How domestic abuse data are captured through the criminal justice system
- Religion and crime in England and Wales
- Child abuse in England and Wales
- Nature of fraud and computer misuse in England and Wales: year ending March 2019

The publications mentioned above are intended only to illustrate the types of reports and findings that are produced from the CSEW. Full details of all publications associated with the CSEW, and crime statistics more generally, can be found on the ONS website⁶.

As well as published reports, anonymised CSEW data is made available through the UK Data Archive at the University of Essex⁷ and through the ONS Secure Research Service⁸. The CSEW is a complex study with

⁴ CSEW Fraud and Cyber-crime Development: Field trial

⁵ <https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/methodologies/improvingcrimestatisticsforenglandandwalesprogressupdate>

⁶ <https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice>

⁷ <https://www.data-archive.ac.uk/>

⁸ <https://www.ons.gov.uk/aboutus/whatwedo/statistics/requestingstatistics/approvedresearcherscheme/>

data organised at different levels (households, individuals, and incidents) and it includes numerous sub-samples who are asked specific questions. Accordingly, considerable effort and expertise is required to analyse the data and to interpret it in a valid manner. Some of the analysis routines that play a key role in the published estimates are implemented after the data have been supplied to the ONS and so are not documented in this report. Further information on how to use the data is available from the UK Data Service⁹.

ONS also produces a User Guide for those interested in understanding CSEW data and outputs which contains further detail on the content and structure of the data¹⁰.

1.3 Structure of the Technical Report

This report documents the technical aspects of the 2019-20 CSEW. The analysis in this report relates to the total sample that was issued in the financial year 2019-20, irrespective of when interviews actually took place. The distinction between issued sample and achieved sample is explained in more detail in Chapter 2 of this report.

The sample design is set out in Chapter 2. Data collection is the major task for the organisation commissioned to conduct the CSEW and forms the central part of this report. Chapter 3 covers the content and development of the questionnaire, while Chapter 4 outlines how the risk ratings system for the 10-15 year olds survey was developed. Chapter 5 details our fieldwork procedure (including response rates, documents and quality control) and Welsh fieldwork. Chapter 6 discusses response rate and reasons for non-response in the core sample. Chapter 7 gives details of the tasks that are involved in preparing the data for analysis, including the coding and offence classification and Chapter 8 covers the preparation and delivery of the CSEW data files. Chapter 9 outlines the weighting required for analysis of the data. Chapter 10 provides the results of some checks on the profile of the CSEW achieved sample against estimates for the population that the CSEW aims to represent.

⁹ <https://www.ukdataservice.ac.uk/>

¹⁰ <https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/methodologies/userguidetocrimestatisticsforenglandandwales>

2. Sample Design

2.1 Introduction

The 2019-20 sample design is similar but not identical to that used for 2018-19.

The key features of the 2019-20 design are as follows:

- A target sample size of 34,500 interviews with adults aged 16 and over who are resident in private households in England and Wales;
- A minimum of 625 interviews per year in each of the 42 police force areas (PFAs)¹¹. This requires a degree of over-sampling in less populous PFAs;
- Use of a bespoke sampling geography for the survey that maximises the heterogeneity of the sample clusters;
- Different levels of sample clustering in different population density segments with every cluster being sampled at least once over a three-year period to create a near un-clustered sample;
- An achieved sample size of up to 3,000 10 to 15-year olds identified through screening at households in which adult interviews have been conducted; and
- Fieldwork conducted on a continuous basis with each sample stratum allocated to a specific quarter in such a way that updated nationally representative estimates are available every three months.

2.2 Sample size and structure

The target sample size for the 2019-20 survey was 34,500 interviews with adults aged 16 and over living in private households in England and Wales. Additionally, the survey had a target of interviewing up to 3,000 10-15 year olds identified through screening within the households that yielded an adult interview.

A minimum of 625 adult interviews were required per police force area (representing a total of 26,250 interviews across 42 PFAs), with the remaining 8,250 adult interviews (to take the total up to 34,500) allocated in proportion among the largest PFAs to maximise the sample efficiency of national estimates. This model provides a national sample efficiency of 95%¹².

The sampling fraction used in each police force area was based on (i) the target sample size and (ii) the observed deadwood¹³ and response rates over the previous survey year. Since these rates are subject to some annual fluctuation at police force area level, the number of addresses to sample in each PFA was inflated by a magnitude of 1.2 to create a pool of reserve addresses. Additionally, it was agreed that within each police force area a range of +/- 50 interviews around the actual target would be deemed acceptable (i.e. for a police force area with a target of 625 achieved interviews, the expected number of interviews should fall in the range 575-675).

Table 2.1 shows the number of addresses anticipated to be required for each police force area at the start of the 2019-20 survey, the actual number of addresses issued, and the target number of interviews. Between

¹¹ For sampling purposes, the City of London police force area is combined with the Metropolitan police force area.

¹² Sample efficiency = effective national sample size due to disproportionate sampling divided by the actual national sample size of 34,500.

¹³ 'Deadwood' addresses are those identified as not being an eligible residential address. The most common type of deadwood is empty or vacant residential properties.

July and December 2019, a total of 827 reserve addresses were issued (c. 137 per month) based on performance in the first quarter of the year to ensure that the annual targets were met.

The actual number of interviews achieved and the final annual response rate for each police force area are shown in [Table 6.6](#).

Table 2.1 Total issued and achieved sample sizes by police force area (2019-20)

Police force area	Anticipated no. of addresses to issue	Actual no. of addresses issued	Target no. of interviews	Target range
Metropolitan/City of London	6,498	6,603	3,899	3,849 – 3,949
Greater Manchester	2,097	2,131	1,410	1,360 – 1,460
Merseyside	1,448	1,471	903	853 - 953
South Yorkshire	1,178	1,194	708	658 - 758
Northumbria	1,109	1,127	781	731 - 831
West Midlands	2,285	2,317	1,355	1,305 – 1,405
West Yorkshire	1,797	1,824	1,162	1,112 – 1,212
Avon & Somerset	1,321	1,340	850	800 - 900
Bedfordshire	957	971	625	575 - 675
Thames Valley	1,758	1,784	1,146	1,096 – 1,196
Cambridgeshire	963	979	625	575 - 675
Cheshire	958	973	625	575 - 675
Cleveland	924	939	625	575 - 675
Devon & Cornwall	1,529	1,551	943	893 - 993
Cumbria	980	995	625	575 - 675
Derbyshire	940	955	625	575 - 675
Dorset	1,044	1,059	625	575 - 675
Durham	846	860	625	575 - 675
Sussex	1,437	1,457	853	803 - 903
Essex	1,544	1,564	903	853 - 953
Gloucestershire	1,012	1,028	625	575 - 675
Hampshire	1,392	1,414	992	942 – 1,042

West Mercia	1,019	1,033	625	575 - 675
Hertfordshire	968	984	625	575 - 675
Humberside	944	960	625	575 - 675
Kent	1,400	1,420	893	843 - 943
Lancashire	1,142	1,160	774	724 - 824
Leicestershire	1,027	1,042	625	575 - 675
Lincolnshire	932	947	625	575 - 675
Norfolk	996	1,011	625	575 - 675
Northamptonshire	971	986	625	575 - 675
North Yorkshire	1,054	1,070	625	575 - 675
Nottinghamshire	956	970	625	575 - 675
Staffordshire	913	929	625	575 - 675
Suffolk	1,053	1,067	625	575 - 675
Surrey	981	996	625	575 - 675
Warwickshire	951	965	625	575 - 675
Wiltshire	996	1,010	625	575 - 675
North Wales	1,017	1,032	625	575 - 675
Dyfed Powys	996	1,011	625	575 - 675
Gwent	869	884	625	575 - 675
South Wales	1,073	1,089	678	628 - 728
TOTAL	54,275	55,102	34,500	

2.3 Sample design

In 2012, Kantar revised the CSEW sample design with the objective of reducing the degree of clustering and so improve the precision of the estimates. To this end, Kantar worked with the mapping experts, *UK Geographics*, to create a set of bespoke, geographically discrete strata for use in the CSEW.

Section 2.3.1 of the 2013-14 Technical Report describes the creation of these strata and they were also the subject of an article in the Survey Methodology Bulletin published by the Office for National Statistics¹⁴. To summarise:

¹⁴ Williams J (2012) The creation of bespoke sample clusters for the Crime Survey for England and Wales 2012-2015, Survey Methodology Bulletin, 71, pp. 45-55

- Every police force area was divided into a set of geographically discrete sample strata, each with an approximately equal number of addresses.
- Each sample stratum was constructed from whole lower level super output areas (LSOAs) so that population statistics could easily be generated for the sample stratum.
- In constructing the sample strata, the design team took account of geographical barriers and the primary road network to ensure that field assignments based upon sample stratum boundaries would be practical.
- The size of each sample stratum was governed by the requirement that approximately 32 addresses should be sampled from each stratum each year.

Each of the 1,639 sample strata is activated¹⁵ once a year and has been allocated to a specific 'activation quarter'. Each activation quarter contains a (stratified) random subsample of the 1,639 sample strata, representative in terms of (i) expected victimisation rates, and (ii) spatial distribution. This minimises the risk of spurious quarter-by-quarter changes in CSEW estimates that are due solely to differences in sample composition.

Once constructed, the 1,639 strata were ranked by the geographical density of addresses within their borders:

- The densest third were classified as belonging to the 'high density segment'
- The least dense third were classified as belonging to the 'low density segment'
- The rest were classified as belonging to the 'mid density segment'^{16 17}

In the 'low density' strata, three geographically discrete sub-divisions were formed (A, B and C), each with an approximately equal number of addresses and constructed from whole LSOAs¹⁸. In the mid density strata, two sub-divisions (A and B) were formed on the same basis. No subdivision was carried out in the high-density strata.

The combination of high-density strata plus the sub-divisions in the mid and low-density strata are termed 'sample units'. Just one sample unit per stratum is used per year following a sequence established in 2012. In most situations, a fieldwork assignment is based on one sample unit¹⁹.

Each survey year has a planned sample unit activation sequence as shown in [Table 2.2](#).

¹⁵ By 'activated' we mean that a sample of addresses is drawn within the stratum, advance letters are sent, and field interviewers start work.

¹⁶ Kantar carried out a small degree of reallocation after this initial classification, essentially to allow a small number of police force areas to obtain the benefits of an unclustered sample over two years rather than three (and every year for the Metropolitan/City police force area).

¹⁷ It should be acknowledged that address density may change over time and that the classification of a stratum as high, mid or low density is specific to 2012.

¹⁸ Stratum subdivisions were designed to be as heterogeneous as possible in terms of crime rates but without forming awkward geographical shapes that would be difficult for interviewers to manage.

¹⁹ Generally speaking, a high-density stratum will contain twice as many addresses as a subdivision within a mid-density stratum and three times as many addresses as a subdivision within a low-density stratum. However, geographically they will be of similar size. Consequently, sample units/fieldwork assignments are roughly equal in size too.

Table 2.2 Sample unit activation in the CSEW (2012-21)

	High density strata	Mid density strata	Low density strata
2012-13	All	'A' subdivisions only	'A' subdivisions only
2013-14	All	'B' subdivisions only	'B' subdivisions only
2014-15	All	'A' subdivisions only	'C' subdivisions only
2015-16	All	'B' subdivisions only	'A' subdivisions only
2016-17	All	'A' subdivisions only	'B' subdivisions only
2017-18	All	'B' subdivisions only	'C' subdivisions only
2018-19	All	'A' subdivisions only	'A' subdivisions only
2019-20	All	'B' subdivisions only	'B' subdivisions only

As noted above, Kantar used a stratified random sampling method to allocate each sample stratum to a specific quarter. This was based upon modelled estimates of the adult victimisation rate using data from the 2008-2011 survey. Four equal sized groups were formed in each PFA based on the modelled victimisation rates.

Additionally, some spatial stratification was carried out to ensure that the allocation per quarter in each PFA had the same broad geographic spread. This was done by using the latitude and longitude values for the 'centroid' address in each sample stratum²⁰. Within each of the four 'victimisation rate' groups in each PFA, the sample strata were sorted by longitude to create three geographic sub-groups (east, central, and west). Finally, the sample strata were ranked by latitude within each of these groups to form a final sequence for systematic allocation.

Although each sample stratum has been allocated to a quarter, they are actually 'activated' on a monthly basis. Consequently, each sample stratum has been randomly allocated a specific month (1,2,3) within its activation quarter. Monthly activation ensures a smooth flow of interviews over time and maximises the representativeness of the datasets, given they are defined by interview date rather than sample activation date. Occasionally, the activation month is switched to improve the management of fieldwork, but activation quarter has remained a fixed characteristic of each sample unit.

Before the 2015-16 survey, the sample strata and their associated sub-divisions were redefined, based on the new LSOAs constructed from 2011 census data rather than 2001 census data. The vast majority of these 2011 LSOAs are identical to a 2001 equivalent and could be allocated to sample strata and associated subdivisions on a simple like-for-like basis. A small number of genuinely new 2011 LSOAs needed to be allocated to sample stratum and subdivision on a spatial 'best fit' basis. This work was carried out by the geographer who had directed the original construction of the sample strata and their associated sub-divisions.

²⁰ The 'centroid' was the most central address in the PSU based on the address distribution rather than on the geographic borders of the sample cluster.

2.4 Sampling of addresses

The Postcode Address File (PAF)²¹ was used as the address source for the CSEW. The PAF is thought to list the addresses for at least 98% of the residential population²². PAF addresses are linked to higher geographies via ONS's National Statistics Postcode Lookup database which is updated four times a year. This database links postcodes to LSOA, allowing addresses to be allocated to sample strata in an unambiguous fashion.

The list of addresses is randomly selected from PAF. The PAF is filtered to exclude obvious non-residential addresses but errs towards over-coverage (i.e. inclusion of addresses that are not yet built or sold or which have been demolished or abandoned). Most of the addresses will be private, residential addresses, but there will always be a proportion of addresses that are not eligible for the survey for one reason or another. These addresses are known as **deadwood** and over the whole sample we might expect anything between 8-12% of all issued addresses to be deadwood. However, this will vary from area to area.

At each address, one of the interviewer's first tasks is to establish whether the address is eligible or not. Addresses that are not traceable, that are non-residential, or that are empty or considered as a second home are all coded as deadwood. Each type of deadwood is considered below.

2.4.1 Non-residential addresses –no private dwelling

Most non-residential addresses are not included on the PAF that is used to select addresses. However, since inclusion is based on the volume of mail a particular address receives, some non-residential addresses with a relatively low volume of mail do get included in the sample.

The most common types of non-residential addresses include factories, businesses, shops, offices, schools, hospitals, churches, etc.

However, an address which may appear non-residential may contain a private residence which shares the same address. For example, a shop may have a flat above it which shares the same address. In this situation the flat would be an eligible residential address. Similarly, a school caretaker may live in a house in the grounds of a school, where the school and the house share the same address. In this situation the caretaker's house would be an eligible residential address.

In both these examples, it is also the case that the shop and the flat or the school and the house actually have slightly different addresses. For example, the shop may be 3 High Street and the flat above it may be 3A High Street. If this is the case, the two properties are treated as completely separate addresses. An interview will only be conducted at the exact address as listed in the sample.

2.4.2 Residential addresses - communal establishments

Another type of deadwood is anything that might be classed as an institution or a communal establishment. Examples include nursing or residential care homes, hotels, hostels, NHS nursing accommodation, college halls of residence, etc. Although these types of addresses are residential, the survey only includes **private** residential addresses.

It is important to distinguish a communal establishment from a private residential establishment. In some cases the distinction between the two can be subtle. Three examples illustrate the potential difficulties:

- While residential care homes for older people are usually classed as communal establishments, sheltered accommodation is generally considered private residential addresses (even where there is a warden)
- While most hostel type establishments are usually classed as communal establishments, bed sits are generally considered to be private residential addresses

²¹ This excludes addresses that receive more than 25 postal items a day.

²² Individuals living in communal accommodation are excluded from the population base.

- While army barracks are usually classified as communal establishments, private residences located on an army base are generally considered to be private residential addresses

In making these distinctions interviewers are instructed to try to think in terms of how people actually live at an address and the extent to which people live independently. Communal living is generally taken to be situations where people share meals together and share communal living space. Where there is a degree of independent living with people generally cooking for themselves or having their own living space this is generally regarded as private residential.

2.4.3 Private residential addresses - vacant or second homes

There are some situations where an address meets the criteria of a private residential address but is not actually occupied. These are probably the most difficult type of addresses to establish positively as deadwood because it is often difficult to make contact with anyone. It can therefore be difficult to establish whether the property is empty or whether the occupants are simply difficult to get hold of.

Addresses are not classed as empty or unoccupied just because an interviewer is unable to make contact with anyone at the address. Either the property must be obviously empty or vacant (e.g. boarded up council flats, properties with no furniture or no sign of occupation) or the interviewer must establish from some other source that no-one is living there. If the interviewer remains unsure about the status of the address, they are instructed to code the outcome as 'Unknown whether address is residential'.

Second homes and holiday homes are another type of residential property that is not eligible for the survey. Again, the main problem with second homes is that it may be difficult to actually make contact with anyone at the address if they are only there occasionally. Therefore, interviewers always try to check with neighbours wherever possible.

In some cases, an individual may be unsure which of their residences should count as their main address and which should count as their second home. If this is the case, they are asked to think about which address they live at for most of the year. This rule only applies if someone has two residences within England and Wales.

2.5 Police Force Area (PFA) sampling

Within each PFA, the number of addresses issued in 2019-20 was based on the target number of interviews to be achieved across the year divided by the estimated address conversion rate. Dividing the total number of addresses issued within each police force area by the overall total number of addresses in the police force area gives a basic address sampling fraction for each PFA.

However, from 2015-16, this basic address sampling fraction was modified within activated sample units to compensate for random variation in the total number of addresses found within each combination of activated sample units.

The revised address sampling fraction for sampling unit x in police force area y in year t is calculated as follows:

$$f_{2xyt} = f_{xyt} * (N_{yt} / (N_{hyt} + 2N_{myt} + 3N_{lyt}))$$

Where:

f_{xyt} = basic year t sample fraction for sampling unit x in police force area y

N_{yt} = total number of addresses in police force area y in year t

N_{hyt} = total number of addresses in high density strata in police force area y in year t

N_{myt} = total number of addresses in *activated* sample units in mid density strata in police force area y in year t

N_{lyt} = total number of addresses in *activated* sample units in low density strata in police force area y in year t

As already mentioned, since conversion rates at PFA level are subject to some fluctuation, it was decided to over sample addresses by a magnitude of 1.2 to create a pool of reserve addresses in each activated sample unit. In the event, 827 reserve sample addresses were issued during the 2019-20 survey year (see table 2.1) and so the PFA address sampling fraction was updated to reflect this slightly larger issued sample.

2.6 Selection of addresses within sample units

In each sample unit, addresses were geographically sorted prior to a systematic sample being drawn using a fixed interval and random start method. Geographic sorting within sample unit was based on LSOA, Output Area, full postcode, and alphanumeric address.

The number of addresses selected for the 2019-20 survey varied within each sample unit but averaged around 39-40. After the addresses had been selected, 20% of addresses were randomly allocated to the reserve sample pool and removed from the main sample. This meant that the average assignment size issued to interviewers was around 33 addresses. In fact, 76% of activated sample units contained between 30 and 36 sampled addresses, 8% had fewer than 30 addresses (minimum 21), and 16% had more than 36 addresses (maximum 47).

2.7 Sampling households and individuals within households

At addresses with more than one dwelling unit, one dwelling unit was randomly selected for interview based on a standard selection algorithm built into the electronic contact script. The number of dwelling units at each address was recorded by interviewers. Within dwellings, very occasionally, interviewers found more than one household resident within a dwelling unit (based on the standard definition of a household). In these cases, one household was selected at random using the same selection process which was used to select a dwelling at multi-dwelling addresses. This additional process for identifying multiple households within dwellings was introduced on the CSEW for the 2015-16 survey.

Within each eligible household one adult was randomly selected for interview based on a standard selection algorithm built into the electronic contact script.

2.8 Sampling of 10 to 15 year olds

The 2019-20 survey had a target of 3,000 interviews with 10-15 year olds identified at the core sampled addresses. Where only one eligible child was identified an interview was always attempted. If more than one eligible child was identified, one child was selected at random to take part in the interview.

3. Questionnaire content and development

3.1 Structure and coverage of the adult questionnaire

The CSEW questionnaire for the adult survey has a complex structure, consisting of a set of core modules asked of the whole sample, a set of modules asked only of random sub-samples, and self-completion modules asked of all 16 to 74 year olds²³. Within some modules there is often further filtering so that some questions are only asked of even smaller sub-samples.

The 2019-20 CSEW questionnaire consisted of the following modules:

1. Household box
2. Perceptions of crime
3. Screener questionnaire
4. Victimization modules for non-fraud incidents identified at the screeners (up to a maximum of six)
5. Victimization modules for fraud incidents identified at the screeners (up to a maximum of six, including the non-fraud incidents)
6. Performance of the Criminal Justice System
7. Mobile phone crime
8. Experiences of the police (Module A)
9. Crime Prevention and Security: Household (Module B)
10. Crime prevention and security: Vehicle Crime (Module C)
11. Crime Prevention and Security: Personal and Online (Module D)
12. Anti-social behaviour
13. Demographics
14. Self-completion module: Drug use and drinking
15. Self-completion module: Gangs and personal security
16. Self-completion module: Domestic violence, sexual victimisation and stalking
17. Self-completion module: Nature of serious sexual assault since age 16

The basic structure of the core questionnaire is shown in Figure 3.1, while the sub-set of respondents who were asked each module of the questionnaire is shown in Table 3.1. The complete questionnaire is documented in Appendix D of Volume 2. This chapter outlines the broad content of each section or module of the questionnaire.

²³ The 2017-18 survey was the first to extend the age range for the self-completion questionnaire from 16-59 to 16-74. This followed an experiment in 2016-17 to test removing the age range altogether (see 2016-17 technical report for details).

Figure 3.1 Flow Diagram of the 2019-20 CSEW Core Questionnaire

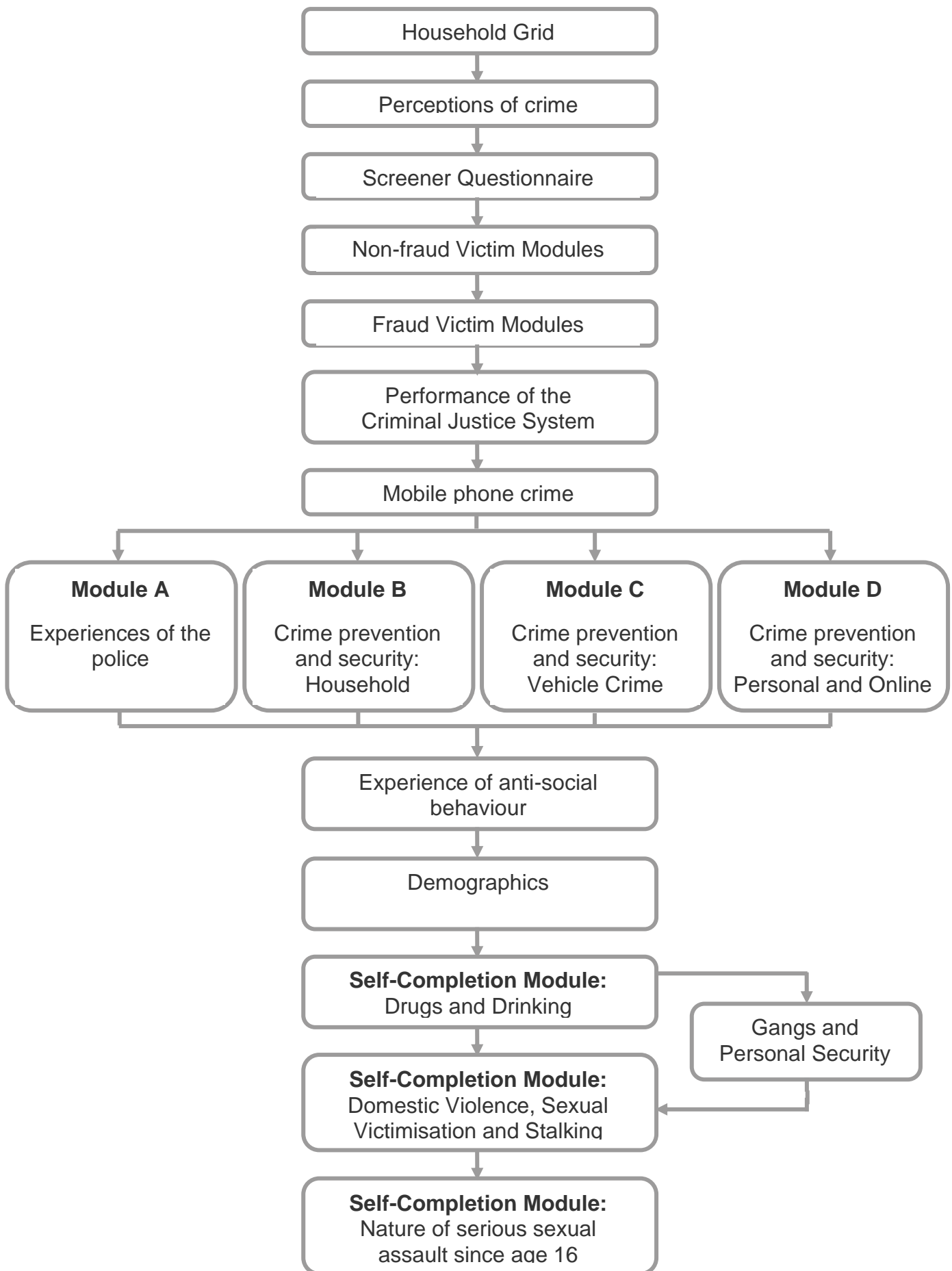


Table 3.1 Modules of the 2019-20 CSEW questionnaire and sub-set of respondents who were asked each module

Questionnaire module	Core sample
Household grid	All
Perceptions of crime	All
Screenener questions	All
Victimisation module	All victims of non-fraud
Fraud victimisation module	All victims of fraud
Performance and experiences of the Criminal Justice System	Random 50% - Groups A and B
Mobile phone crime	All
Module A: Experiences of the police	Random 25% - Group A
Module B: Crime prevention and security: Household	Random 25% - Group B
Module C: Crime prevention and security: Vehicle Crime	Random 25% - Group C
Module D: Crime prevention and security: Personal and online	Random 25% - Group D
Anti-social behaviour	Random 75% - Groups B, C and D
Demographics	All
Self-completion module: Drugs and drinking	All aged 16-74
Self-completion module: Gangs and Personal Security (16-29 year olds only)	Random 50% Groups A and B aged 16-29 years old
Self-completion module: Domestic violence, sexual victimisation and stalking	All aged 16-74
Self-completion module: Nature of serious sexual assault since age 16	All victims of serious sexual assault

Allocation of respondents to each part-sample module was done using an algorithm based on the pre-allocated serial number of each address.

Almost every question in the survey had a don't know and refused option that the interviewer could use. At most questions these options did not appear on the screen to try to ensure that interviewers did not overuse them. Similarly, on most questions with show cards, the don't know and refused options were not presented to respondents, meaning the respondent had to spontaneously mention these responses.

In the questionnaire in Appendix D of Volume 2, don't know and refused codes are only shown if they were explicit response categories and so actually appeared as an option on the screen or show card.

3.1.1 Household grid

Basic socio-demographic details (age, sex, marital status, relationship to respondent) were collected in the household grid for every adult in the household. Additionally, the age, sex and relationship to the respondent of all children under 16 years old were also collected.

The household grid was also used to establish the Household Reference Person (HRP)²⁴ which is the standard classification used on all government surveys and is based on the following criteria:

- The HRP is the member of the household in whose name the accommodation is owned or rented. In households with a sole householder that person is the HRP.
- In households with joint householders the person with the highest income is taken as the HRP.
- If both householders have exactly the same income, the older is taken as the HRP.
- Finally, at the end of the household grid there are some factual questions around length of time living at the address, internet access, and vehicle ownership, all of which are required for filtering of the screener questions.

3.1.2 Perceptions of crime

The household grid was followed by a series of attitudinal and behavioural questions around particular aspects of crime and anti-social behaviour. Questions included:

- Impact of crime on quality of life (Module D respondents only)
- Perceptions of personal safety (Module D respondents only)
- Worries about being a victim of different types of crime (Module B, C and D respondents only);
- Perceptions of anti-social behaviour in the local area (Module A respondents only)
- Perceptions of national and local crime rates and how these have changed (Module A, B and C respondents only)
- Behaviour in relation to going out and frequency of visiting certain places (pubs or bars)

3.1.3 Screener questions – Non-fraud

All respondents were asked whether they had experienced certain types of crimes or incidents within a specified reference period, namely the last 12 months from the date of interview.

Questions were designed to ensure that all incidents of crime within the scope of the CSEW, including relatively minor ones, were mentioned. The screener questions deliberately avoided using terms such as 'burglary', 'robbery', or 'assault', all of which have a precise definition that respondents might not know or fully understand the precise meaning. The wording of these screener questions has been kept consistent

²⁴ Prior to 2001 all previous surveys collected details of the Head of Household.

since the CSEW began to ensure comparability across years, apart from some minor updating of some terminology.

To try and encourage respondents to recall events accurately, a life event calendar was offered to all respondents to act as a visual prompt when answering the screener questions. The idea was to try and place events or incidents in some sort of meaningful context for each respondent by building up a picture of events that have happened to them in the last year (e.g. birthdays, anniversaries, holidays, starting a new job, etc.) that are memorable.

Appendix F in Volume 2 has an example of the calendar used on the 2019-20 core adult survey.

Depending upon individual circumstances, a maximum of 25 screener questions were asked which can be grouped into four main categories:

- All respondents who owned vehicles or bicycles were asked about experience of vehicle-related crimes (e.g. theft of vehicles, theft from vehicles, damage to vehicles, bicycle theft);
- All respondents were asked about experience of property-related crimes in their current residence (e.g. whether the property was broken into, whether anything was stolen from the property, whether the property was damaged);
- All respondents who had moved in the last 12 months were also asked about their experience of property-related crimes at their previous residence(s); and
- All respondents were asked about experience of personal crimes (e.g. whether any personal property was stolen, whether any personal property was damaged, whether they had been a victim of violence or threats)

The questions are designed to ensure that the respondent does not mention the same incident more than once. As a check, at the end of the screener questions, the interviewer is shown a list of all incidents recorded and asked to check with the respondent that all incidents have been recorded and nothing has been counted twice. If there is any evidence of double counting the respondent has an opportunity to correct the information before proceeding.

Within the screener questions there is a crucial distinction between **household** incidents and **personal** incidents.

All vehicle-related and property-related crimes are counted as household incidents. Respondents are asked whether anyone currently residing in their household has experienced any relevant incidents within the reference period. A typical example of a household incident is criminal damage to a car. It is assumed that the respondent will be able to recall these incidents and provide information even in cases where he/she was not the owner or user of the car.

Personal incidents refer to all crimes against the individual and so only relate to things that have happened to the respondent personally, but not to other people in the household. This is often a difficult concept for respondents to understand as their natural inclination is to tell the interviewer about incidents affecting other members of their household. An example of a personal incident would be an assault. An assault against other household members (no matter how serious) should not be recorded, unless the respondent was also assaulted as part of the same incident.

3.1.4 Screener questions – Fraud

From October 2015, new screener questions covering experiences of fraud and cybercrime during the previous 12 months were included. The fraud screener questions were asked to all respondents and were administered in the same way as the traditional non-fraud screeners.

The six main topic areas covered by the fraud screeners were:

- Incidents which occurred as a direct result of a previous non-fraud crime
- Personal information or account details been used to obtain money, or buy goods or services without permission
- Being tricked or deceived out of money or goods
- Attempts to trick or deceive out of money or goods
- Theft of personal information or details held on your computer or in on-line accounts
- Computer or other internet-enabled device being infected or interfered with by a virus

3.1.5 Victimisation modules

All incidents identified at the screener questions are followed through in more detail in the victimisation module. Incidents are covered in a specific priority order which has been kept consistent since the start of the CSEW.

Identification and ordering of incidents for victimisation modules

In 2019-20, 77% of adult respondents did not report any incidents of crime during the last 12 months, meaning that no victimisation modules were completed as part of the interview.

Where a respondent had experienced one or more incidents in the reference period, the questionnaire script automatically identified the order in which the modules were asked. This prioritisation process was adjusted to take account of fraud when these screeners were added to the survey. Fraud crimes were given a lower priority than the existing non-fraud crime types. The automatic selection meant that the interviewer had no discretion about the selection or order of the modules²⁵. The priority ordering used by the script was as follows:

- **According to the type of crime.** Non-fraud victimisation modules were asked first, in reverse order to the screener questions. Broadly speaking this means that all personal incidents were asked before property-related incidents, which were asked before vehicle-related incidents. Fraud victimisation modules were asked but in the same order as the fraud screener questions. Overall, across both non-fraud and fraud crimes a maximum of six victimisation modules were completed, with non-fraud incidents taking priority.
- **Chronologically within each type of crime.** If a respondent reported more than one incident of the same type of crime, modules were asked about the most recent incident first and worked backwards chronologically.

²⁵ In the case of the incidents of sexual victimisation or domestic violence, the interviewer had an option to suspend the victimisation module, as this might make the respondents feel uncomfortable or endanger the respondent in some situations. The interviewer would then attempt to arrange a revisit at a time that would be more convenient, for example when other household members would not be present.

If six or fewer incidents were identified at the screener questions, a victimisation module was completed for all of the incidents reported. For non-fraud cases, the first three modules contained a set of detailed questions relating to each incident (called 'long' modules). The second three modules contained a sub-set of key questions (called 'short' modules) which would still allow the incidents to be classified. This approach was done to minimise respondent burden by limiting overall interview length. Fraud and computer misuse victimisation modules included a different set of questions which were asked for every fraud or computer misuse incident (i.e. no distinction between long and short modules).

In the 2019-20 survey, a total of 10,252 victimisation modules were completed by 7,248 individual victims, with 22.6% of all respondents reporting at least one incident (see Table 3.2).

Table 3.2 Core sample respondents who completed victimisation modules, 2019-20 CSEW

	N	% of all respondents	% of victims
Non victims	24,780	77.4	
Victims²⁶	7,248	22.6	
No. of victim modules completed			
1	5,320	16.6	73.4
2	1,307	4.1	18.0
3	359	1.1	5.0
4	138	0.4	1.9
5	55	0.2	0.8
6	69	0.2	1.0
Total	10,252		
<i>Bases:</i>		32,028	7,248

Defining a series of incidents

Most incidents reported represent one-off crimes or single incidents. However, in a minority of cases a respondent may have been victimised a number of times in succession. At each screener question where a respondent reported an incident, they were asked how many incidents of the given type had occurred during the reference period. If more than one incident was reported, the respondent was asked whether they thought that these incidents represented a 'series' or not. A series was defined as "the same thing, done under the same circumstances and probably by the same people". Where this was the case, only one victimisation module was completed in relation to the most recent incident in the series. Again, this was done to minimise respondent burden.

²⁶ Victims refers to the number of respondents who started at least one victimisation module. This is slightly different to the number of respondents who reported at least one incident at the screen questions (n=7,394).

In fraud cases the definition of a series is more complex, as the survey is intended to replicate the way in which the police would record fraud incidents as close as possible. The key measures for identifying a series with fraud offences is whether all the incidents are identified at the same time, and whether the victim responded in the same way. This is designed to ensure that cases of fraud involving multiple transactions on a single account are counted as a single incident rather than multiple incidents. For example; if someone discovers four separate transactions on their bank account these will be recorded as a single incident rather than four separate incidents or a series. However, if they later discover more transactions on their account then this would be recorded as a separate incident or as the second incident in a series.

There are two practical advantages to the approach of only asking about the most recent incident where a series of similar incidents has occurred. First, since some (although not all) incidents classified as a series can be petty or minor incidents (e.g. vandalism) it avoids the need to ask the same questions to a respondent several times over. And second, it avoids using up the limit of six victimisation modules on incidents which may be fairly trivial, while missing out potentially more serious incidents.

In 2018-19, 88% of all victimisation modules related to single incidents and 12% related to a series of incidents. This split between single and series incidents was broadly the same as on previous surveys.

In the rare cases where a respondent has experienced a mixture of single incidents and a series of incidents the interview program has a complex routine which handles the sequence of individual and series incidents and allows the priority ordering of the victimisation modules to be decided.

In terms of estimating the victimisation rates, series incidents receive a weight corresponding to the number of incidents in the series that fall within the reference period, subject to a maximum limit that is specific to the offence code group ([see section 9.7](#)). This is a relatively recent change to how the data is weighted as previously all offence types were capped at a limit of five.

Content of victimisation module

The victimisation module collects the key information needed to classify each incident to a particular offence type, which is the basis for calculating the prevalence and incidence rates. It contains three types of information:

- **The exact month(s) in which the incident or series of incidents occurred.** In a few cases, respondents may have reported an incident which later turns out to have been outside the reference period. In such cases, the victimisation module is simply by-passed. If respondents were unsure about the exact month in which something happened, they were asked to narrow it down to a specific quarter. For incidents that were part of a series, respondents were asked how many incidents occurred in each quarter and the month in which the most recent incident had occurred.
- **An open-ended description of the incident where the respondent describes exactly what happened in their own words.** The open-ended description is vital to the accurate coding of offences that takes place in the office. Short, ambiguous or inconsistent descriptions can often make offence coding difficult. In fraud victimisation modules a second open-ended description is included to collect information about the action the respondent took following the fraud or attempted fraud, as this is a key aspect of the fraud offence coding. At the end of each victimisation module, the original open-ended description that the interviewer had entered at the start is re-capped, along with the answers to some of the key pre-coded questions. By presenting this information on a single screen, interviewers have the chance to confirm with respondents that the information is correct and consistent. If the respondent and/or interviewer wish to add or clarify any information they can do this.

- **A series of key questions used to establish important characteristics about the incident.**
These include where and when the incident took place; whether anything was stolen or damaged and, if so, what; the costs of things stolen or damaged; any details of the offenders (if known); whether force or violence was used and, if so, the nature of the force used and any injuries sustained; and whether the police were informed or not. While many of the questions in the fraud victimisation module reflect the non-fraud module there are also other questions which are more relevant for these specific types of crime.

3.1.6 Reference dates

In the questionnaire script reference dates were automatically calculated based on the date of interview and appropriate text substitution was used to ensure that the questions always referred to the correct reference period.

Because the 12-month reference period changed each month throughout the fieldwork year, some date-related questions in the victimisation module had different text each month to reflect this changing reference period. Thus, for example, any interviews conducted in July 2019 would use the reference period “*since the first of July 2018*”. This means that in practice the 12-month reference period consisted of the last 12 full calendar months, plus the current month (i.e. slightly more than 12 months). This is taken into account when the victimisation rates are estimated.

In the previous section it was noted that for each incident the respondent is asked which month of the year the incident happened in. At these questions the code frame presented to the interviewer and respondent always displays the last 13 months counting back from the date of interview.

If respondents are unable to narrow it down to a particular month, they are then asked for the quarter of the year it happened in. Additionally, where respondents have reported a series of incidents in the last 12 months, they are asked how many incidents happened in each quarter. The time period used for both these questions is not ‘rolling quarters’ but rather are fixed to match the standard quarters used in both the survey design and in terms of how the estimates are reported (i.e. January – March, April – June, July – September, October – December).

Since the reference period is based on a rolling 12 months based on the month of interview it is important in cases where only the quarter is recorded to be able to establish whether the incident is in scope (within the last 12 months) or out of scope (more than 12 months ago). This requires some questions within the victimisation module to have an adjusted code frame which differs based on the exact month of interview. This is illustrated in Table 3.3 below for the full year 2019-20. In each case the first code is always out of scope (more than 12 months ago) and the other codes are in scope.

Table 3.3 Code frame by month of interview at the victimisation module

<p>Interview month=April 2019</p> <ol style="list-style-type: none"> 1. Before the 1st of April 2018 (Out of scope) 2. Between April and June 2018 3. Between July and September 2018 4. Between October and December 2018 5. Between January and March 2019 6. Between the 1st of April 2019 and present <p>Interview month=May 2019</p> <ol style="list-style-type: none"> 1. Before the 1st of May 2018 (Out of scope) 2. In May or June 2018 3. Between July and September 2018 4. Between October and December 2018 5. Between January and March 2019 6. Between the 1st of April 2019 and present <p>Interview month=June 2019</p> <ol style="list-style-type: none"> 1. Before the first of June 2018 (Out of scope) 2. In June 2018 3. Between July and September 2018 4. Between October and December 2018 5. Between January and March 2019 6. Between the 1st of April 2019 and the present <p>Interview month =July 2019</p> <ol style="list-style-type: none"> 1. Before the 1st of July 2018 (Out of scope) 2. Between July and September 2018 3. Between October and December 2018 4. Between January and March 2019 5. Between April and June 2019 6. Between the 1st of July 2019 and present <p>Interview month=August 2019</p> <ol style="list-style-type: none"> 1. Before the 1st of August 2018 (Out of scope) 2. In August or September 2018 3. Between October and December 2018 4. Between January and March 2019 5. Between April and June 2019 6. Between the 1st of July 2019 and present <p>Interview month=September 2019</p> <ol style="list-style-type: none"> 1. Before the 1st of September 2018 (Out of scope) 2. In September 2018 3. Between October and December 2018 4. Between January and March 2019 5. Between April and June 2019 6. Between the 1st of July 2019 and present 	<p>Interview month=October 2019</p> <ol style="list-style-type: none"> 1. Before the 1st of October 2018 (Out of scope) 2. Between October and December 2018 3. Between January and March 2019 4. Between April and June 2019 5. Between July and September 2019 6. Between the 1st October 2019 and present <p>Interview month=November 2019</p> <ol style="list-style-type: none"> 1. Before the 1st of November 2018 (Out of scope) 2. In November or December 2018 3. Between January and March 2019 4. Between April and June 2019 5. Between July and September 2019 6. Between the 1st of October 2019 and present <p>Interview month=December 2019</p> <ol style="list-style-type: none"> 1. Before the 1st of December 2018 (Out of scope) 2. In December 2018 3. Between January and March 2019 4. Between April and June 2019 5. Between July and September 2019 6. Between the 1st of October 2019 and present <p>Interview month=January 2020</p> <ol style="list-style-type: none"> 1. Before the 1st of January 2019 (Out of scope) 2. Between January and March 2019 3. Between April and June 2019 4. Between July and September 2019 5. Between October and December 2019 6. Between the 1st of January 2020 and present <p>Interview month=February 2020</p> <ol style="list-style-type: none"> 1. Before the 1st of February 2019 (Out of scope) 2. In February or March 2019 3. Between April and June 2019 4. Between July and September 2019 5. Between October and December 2019 6. Between the 1st of January 2020 and present <p>Interview month =March 2020</p> <ol style="list-style-type: none"> 1. Before the 1st of March 2019 (Out of scope) 2. In March 2019 3. Between April and June 2019 4. Between July and September 2019 5. Between October and December 2019 6. Between 1st of January 2020 and present
---	---

3.1.7 Performance and experiences of the Criminal Justice System

A random sub-set of respondents were asked their perceptions about the effectiveness and fairness of both the Criminal Justice System (CJS) as a whole, as well as about the individual agencies that make up the CJS (the police, the courts, the CPS, the probation service and prisons).

A second set of questions were asked to all respondents about their levels of trust and confidence in the police, both nationally and locally. Questions covered overall trust in the police as an institution, perceptions of how good a job the local police are doing, and questions related to specific aspects of local policing.

A third set of questions related to respondents' knowledge of Police Crime Commissioners, whether they had contacted them and how likely they would be to contact their local Police Crime Commissioner.

Finally, a set of questions asked respondents about their actual experiences of the court system and if they had attended any type of court in the last 12 months in any capacity and, if so, how they rated their experiences.

3.1.8 Mobile phone crime

Although mobile phones stolen from the respondent should be identified in the victimisation module, thefts from other members of the household are not covered. Consequently, in this module, all respondents were asked if anyone else in the household had had a mobile phone stolen in the last 12 months and, if so, from whom the phone had been stolen. Respondents were asked to include incidents where mobile phones stolen had been stolen from children in the household.

3.1.9 Sub-sample modules (A-D)

Respondents were randomly allocated to one of four modules (A-D). The random allocation maintains a representative sub-sample in each of the modules and ensure roughly equal sample sizes.

Module A: Experiences of the police

Topics covered in this module included:

- the extent to which police are visible in the local area
- what sources they use to get information about their local police
- whether they have had contact with their local police and, if so, for what reasons and in what form did the contact take; and
- whether respondents have made a complaint about the police and, if so, how they felt their complaint was dealt with

Module B: Household crime prevention and security

Topics covered in this module included:

- whether or not respondents have a range of security measures installed at their home
- whether respondents have changed their home security measures or behaviour in the last 12 months

Module C: Vehicle crime prevention and security

Topics covered in this module included:

- whether or not respondents have a range of security measures on their vehicles

Module D: Personal and online crime prevention and security

Topics covered in this module included:

- personal security habits and the steps that respondents take to reduce their chances of being a victim of crime when they are out and about in public
- whether or not respondents take measures to keep themselves safe online

3.1.10 Anti-social behaviour

This module included:

- perceptions of anti-social behaviour in the local area
- experiences of different types of anti-social behaviour

In previous years a much longer module of questions was asked around anti-social behaviour, but this was cut back for the 2019-20 survey to a few key measures.

3.1.11 Demographics

This section collected additional information on the respondent and the Household Reference Person (where this was not the same as the respondent). Question topics included:

- health and disability
- employment details²⁷
- ethnicity and national identity
- educational attainment and qualifications
- housing tenure
- household income.

Some more sensitive questions related to wellbeing, sexual orientation and gender identity were included in the self-completion module of the survey (see next section) and so were only asked of 16-74 year olds.

3.1.12 Self-completion modules

The self-completion modules were asked of respondents aged 16 to 74 years of age. These modules are all presented as computer assisted self-completion (CASI) modules to ensure respondent confidentiality in answering these questions.

The respondent was asked to follow the instructions on the laptop screen and enter their answers accordingly. Practice questions were included before the start of the self-completion module to give the interviewer an opportunity to show the respondent the different functions of the tablet. If the respondent was unable or unwilling to complete the modules using the computer, the interviewer could administer the self-completion. In these circumstances, respondents were only asked the modules on drug use and drinking but not the module on domestic violence, sexual assault and stalking. Interviewer assistance and the presence of others while completing these modules was recorded by the interviewer ([see Table 5.3](#)).

In 2016-17, Kantar experimented with increasing the age limit on the self-completion module, from 59 years of age to 74 years. Results showed that adults of this age were able to successfully answer self-completion questions, although refusal rates were slightly higher compared with other age groups and respondents were more likely to require help from an interviewer. From 2018-19 the upper age limit for the self-completion modules was permanently raised to 74 years.

Self-completion module – illicit drug use and alcohol consumption

All core respondents were asked this series of questions on drug and alcohol use. The module covered a wide variety of drugs of all classes including amphetamines, cannabis, cocaine, ecstasy, heroin, and many others. Respondents were asked whether they had ever taken each drug and, if so, whether they had taken

²⁷ Where the respondent was not the Household Reference person occupation details were also collected about the HRP.

it in the last 12 months and whether they had taken it in the last month. The list of drugs included a drug that did not exist (Semeron) to attempt to identify instances of over reporting.

Respondents were also asked about taking psychoactive substances such as nitrous oxide and substance formerly known as legal highs and any prescription-only painkillers that were not prescribed in the last 12 months.

Finally, respondents were asked about their alcohol consumption, including how often they had drunk alcohol in the past 12 months, how often they had felt drunk and whether they thought they had driven a vehicle when they were over the legal alcohol limit.

Gangs and Personal Security

Respondents aged 16-29 years old were routed to an additional short self-completion module containing questions on street gangs and knife carrying.

Domestic violence, sexual victimisation and stalking

All respondents aged 16-74 who accepted the self-completion module without any interviewer assistance were routed to a self-completion module covering experiences of domestic violence, sexual victimisation and stalking.

The module was largely based on questions first developed in 2001 (and modified in 2004-05) to measure the prevalence of domestic violence, sexual victimisation, and stalking. However, in October 2015, the law was changed to make coercive and controlling behaviour an offence and the questions were updated to reflect this change. Following a review of these questions they were re-developed to help improve respondent understanding. In 2017-18 a split sample experiment was conducted to test two different variants of questions and to compare the prevalence estimates produced by each. Based on the findings of this experiment the original questions were modified slightly.

This current set of questions on inter-personal violence cover the following topics:

- experience of domestic violence by either a partner or by another family member since age 16 and in the last 12 months
- experience of less serious sexual assault since age 16 and in the last 12 months
- experience of serious sexual assault since age 16 and in the last 12 months
- experience of stalking since age 16 and in the last 12 months

Nature of serious sexual assault since age 16

Those who had been subjected to serious sexual assault since the age of 16 were asked supplementary questions about the nature of the sexual assault. The questions covered:

- frequency of incidents;
- whether the police came to know or not;
- whether drugs or alcohol were involved;
- whether respondent suffered any injuries or sought any medical help; and
- whether respondent had to take any time off work

3.2 Question development and testing

In most survey years question testing is a standard component of the CSEW questionnaire development process. This usually takes the form of cognitive testing conducted via face-to-face interviews. However, for the 2019-20 survey only a small number of questions required testing and so it was decided not to conduct any cognitive testing. Instead, an online pilot approach was used.

The questions to be tested were included as part of a short online survey, along with some open text questions to gather qualitative feedback from respondents about the new questions. The survey was run on Kantar's Profiles Panel with a sample size of 1,000 respondents. Quotas were set to ensure a nationally representative sample was achieved. Alternative versions of the same question were tested in some cases, with each version being asked of a random sample of 500.

3.3 Structure and coverage of the 10-to-15 year-olds survey

An extensive development and testing phase was undertaken prior to the launch of the 10 to 15 year olds survey. The results of this phase are documented in the development report published in 2010.²⁸

The 2019-20 CSEW questionnaire for 10 to 15 year olds covered:

- Schooling
- Crime screener questions – personal incidents only
- Victimisation module
- Self-completion module; and
 - Use of the internet
 - Attitudes to online safety and risk perception
 - Bullying
 - Speaking to or meeting strangers online
 - Sending and receiving messages
 - Online security
 - Street gangs
 - Opinions on burglary and violence
 - School truancy
 - Carrying knives
 - Drinking behaviour and cannabis use
- Demographics

3.3.1 Schooling

This module included questions about whether the respondent attended school and, if so, what school year they were in (school year is used later in the questionnaire to help respondents recall exactly when incidents of crime took place).

3.3.2 Crime screener questions

All respondents were asked whether they had experienced certain types of crimes or incidents within the last 12 months. To aid recall respondents were given a life events calendar similar to the one used on the adult survey. Appendix F in Volume 2 has an example of the calendar used on the 10 to 15 year olds survey.

Respondents in the 10 to 15 year-olds survey were not asked about household incidents as these would have been covered in the interview with the adult household member. Respondents were asked:

- Whether anything had been stolen from them;
- Whether anyone had attempted to steal something from them;
- Whether anyone had deliberately damaged their property;
- Whether anyone had deliberately kicked, hit, pushed or been physically violent towards them in any other way; and
- Whether they had been threatened

²⁸ Extending the British Crime Survey to children: a report on the methodological and development work.

3.3.3 Victimization modules

All incidents identified at the screener questions were followed up in more detail in the victimisation module. Incidents were covered in specific priority order up to a maximum of three:

- according to the type of crime;
- chronologically within each type of crime – if a respondent reported more than one type of incident of the same crime type, victim modules were asked about the most recent incident first and worked backwards chronologically; and
- up to a maximum of three full victim forms

As with the core survey the victimisation module collected the key information required for classification of offences including:

- the exact month in which the most recent incident took place;
- an open-ended description of the incident; and
- a series of key questions to establish important characteristics of the incident

3.3.4 Self-completion modules

Several modules contained potentially sensitive questions and were therefore included in the self-completion section of the survey. As in the core survey, practice questions were included so that the interviewer could explain to the respondent how to use the computer.

Use of the internet - Respondents were asked whether they had used the internet in the last 12 months and, if so, what they used the internet for.

Attitudes (feeling of safety and risk perception) – Respondents were asked questions about their attitudes towards their safety and perception of risks, including how safe they feel walking around their area during the day, their general confidence levels, if their parents/ guardians know what they're doing when they are out, if they avoid talking to strangers, and their worries about sharing personal information online.

Bullying – This module asked whether the respondent had been bullied either offline or online and, if this was the case, some follow up questions were asked about the nature and extent of the bullying.

Speaking/Meeting strangers online – This module asked respondents about the contact they had had with people online on any platform (e.g., on a computer, mobile phone or gaming console). Respondents were asked who they spoke to online, whether they knew them or not, methods of communication and whether they had met anyone in person as a result of an online exchange. If the respondent had met up with a stranger in person, they were asked about this meeting, including whether they had told anyone in advance (e.g. a parent, friend).

Sending and receiving messages – This module asked respondents whether they had sent or received any sexual messages in the last 12 months. If they had, the respondent was asked what type of messages, how these were sent/received, if they were bothered about it, and if they told anyone. If the respondent sent a sexual message to someone, they were also asked if anyone still has the message saved, if it was posted or shown to anyone else without agreement, and how they felt about it.

Online security – Respondents were asked questions about how their parents monitored what they did online, including if they had any rules or restrictions about what they could do and if their parents talked to them about online safety. They were also asked about if they'd received information about keeping safe online and where they got their online safety information from.

Street gangs – Respondents were asked how much of a problem they believed street gangs to be in their local area. They were also asked whether they knew anyone who was a member of a street gang and whether they themselves were a member of a street gang.

Opinions on burglary and violence – Two questions were asked about how ‘wrong’ the respondent thought it was to break into a building to steal something or to use a weapon or force to get money/things from another young person.

School truancy – Three questions were asked covering whether the respondent had missed school without permission in the preceding 12 months, how many times they had missed school without permission and whether they had been suspended or excluded from school.

Carrying knives – Two questions were asked about whether the respondent knew anyone who carried a knife for their own protection and if they had carried a knife themselves for protection in the last 12 months.

Drinking behaviour – This section of questions asked whether the respondent had ever drunk alcohol, whether they had ever been drunk, and how often they had been drunk.

Cannabis use – Respondents were asked whether they had ever tried cannabis, and how often they had tried it.

3.3.5 Demographics module

The demographics module included questions regarding ethnicity, religion and whether the respondent had a disability or suffered from a long-term illness.

4. Developing a risk rating system

This chapter discusses the risk scoring and rating system implemented alongside the cybercrime and online behaviours module implemented in the 10-15 year olds survey in April 2019.

4.1.1 Background

In 2016, ONS and Kantar began development for a new cybercrime and online behaviour module for the 10-15-year olds survey. This module was developed to measure the extent of victimisation of cyber related crime among children. Cybercrime in this context is defined as any crimes facilitated by technology and/or the internet, included both cyber enabled and cyber dependent crime. The module includes questions about online bullying, speaking to and meeting up with strangers, sending and receiving messages or images of a sexual nature, and attitudes towards staying safe online.

In 2018, the module was reviewed by the ONS Ethics Committee, with input from the NSPCC. During this process it was suggested that a mechanism to inform parents and children of the potential risk of the child's online behaviour should be developed. As a result, a risk score was devised based on responses to key questions throughout the survey. This risk score was in turn used to develop a risk rating for each section of the module. The highest rating across all sections was then taken as the overall risk rating for the respondent: either 'low risk', 'medium risk', or 'high risk'. The scoring system was developed through collaboration between the ONS, NSPCC and Kantar. The scores are weighted depending on the level of risk represented by each specific response. Aggravating behaviours serve to multiply and increase risk scores.

4.1.2 Pilot Study

In 2018, a small-scale pilot was conducted with parents and children aged 10-15 to explore reactions to the risk rating and understanding of the score among both parents and 10-15-year olds. Interviews were conducted with parents and children aged between 10 and 15. Each interview consisted of three parts: an initial interview with the parent, an interview with the child to complete the survey and a final interview with the parent to discuss the risk rating.

The pilot had three aims:

1. To understand how parents and children interpreted the information about the survey and the risk rating.
2. To understand whether the risk rating and confidentiality statements affect children's responses to the self-completion module (specifically, whether they select different responses as a result)
3. To understand the reaction of parents and children to the risk rating and survey materials.

Findings from the pilot study determined that, in principle, a risk rating system could be implemented from April 2019. Following the pilot study, the matrix scoring system was updated to reflect questionnaire changes and further input from ONS and NSPCC.

4.1.3 Calculating risk scores and ratings

Each section in the cybercrime and online behaviour module have their own scoring and risk rating system²⁹. The highest rating in any of the sections was taken as the overall rating.

A matrix was used to assign risk values to responses at key questions or combinations of questions in each of these sections. The questions that were used can be found in Appendix C of Volume 2. A response code could either increase the score by a fixed value, multiply the score, or flag a case as high risk (so overriding the score). These actions could be assigned to individual response codes or in some cases to combinations of responses which indicated a level of risk. For example, if a respondent indicates that they are being bullied every day, and they have not reported it to anyone.

As each section has a separate scoring system, each section has a separate rating system to account for the differences in score distributions (see Table 4.1). These rating brackets were developed in collaboration with the ONS and NSPCC and determined by examining potential responses and the scenarios they may indicate.

Table 4.1 Module rating systems

Section	Low rating	Medium rating	High rating	Max score
Bullying	0-9	10-17	18-48	48
Speaking to strangers	0-7	8-22	23-46	46
Meeting strangers	0-8	9-21	22-1920	1920
Receiving images	0-8	9-18	19-50	50
Sending images	0-9	10-18	19-154	154

Risk scores and ratings were calculated monthly with SPSS syntax using an export of the live data.

4.1.4 Informing respondents and parents/guardians of risk ratings

Letters detailing the participants risk rating were sent to both the 10-15-year-old and the parent/guardian who provided consent for the interview. Both letters provided the risk rating produced for the child and a short text providing general examples of what this level of risk meant. The letter also contained contact information for sources of information for parents to support their children staying safe online.

The score, which section the rating was taken from, or details of what the child reported in the survey were not included in the letter. The letter addressed to the 10-15-year-old participant contained the same information but in a simpler and child-friendly manner and provided details for age appropriate support to staying safe online. Letters were despatched monthly.

Copies of the risk rating letters sent to both parents and children can be found in Appendix C of Volume 2.

²⁹ Sections are bullying, speaking to strangers online, meeting strangers, receiving images of a sexual nature and sending images of a sexual nature.

5. Fieldwork

This chapter documents all aspects of the data collection process, focusing on fieldwork procedures, the management of fieldwork across the survey year, quality control procedures and response rates achieved across the different samples.

5.1 Briefing of interviewers

All new interviewers working on the Crime Survey for England and Wales are required to attend a full day face-to-face briefing before they can work on the survey. This initial briefing is followed by a half day follow up briefing around 4-6 months later to review interviewer's progress on their early assignments.

In 2019-20, six full day interviewer briefings were held with a total of 80 interviewers attending. A further full day briefing, scheduled for April, was cancelled due to the coronavirus pandemic.

In previous years, interviewers working on the CSEW have been required to attend a half day refresher briefing annually. From 2018-19, this requirement was reduced to a biennial refresher briefing. There were nine half-day refresher briefings held, attended by 130 interviewers. These were held in February and March 2020. A further six refresher briefings, scheduled for April, were cancelled due to the coronavirus pandemic. The cancelled refresher briefings were replaced with 10 WebEx refresher sessions, which were attended by 235 interviewers.

5.2 Supervision and quality control

Several methods were used to ensure the quality and validity of the data collection operation.

A total of 167 CSEW assignments, approximately 10% of all CSEW assignments allocated in 2019-20, were supervised. Assignments supervised tended to be those assigned to less experienced interviewers. Interviewers new to random probability sample surveys were also accompanied on the first day of their CSEW assignment by a supervisor.

A total of 4,135 addresses across 610 separate CSEW assignments were validated during the year; 13% of all addresses where an interview was achieved. Validation was carried out mainly by telephone. Where no telephone number was available, a short postal questionnaire was sent to the address to collect the same information.

Addresses for validation were selected on the basis of Kantar's standard field quality procedures, where all interviewers have their work checked at least twice a year. For these checks, full assignments were validated (i.e. all addresses in the assignment where an interview was achieved). On top of these whole assignment checks a random 5% of all CSEW interviews were also validated.

In addition to validation, the performance of all interviewers working on the survey was monitored closely. Where an underperforming interviewer was identified they were offered additional training and, in some cases, accompanied by an experienced supervisor on their next assignment.

To ensure that the data collected was robust and collected in a consistent manner, Kantar also conducted systematic quality checks on the data on a quarterly basis at an interviewer level. This involved collating

responses across several key indicators at an interviewer level (e.g. average interview length, number of victims identified, take up of the self-completion modules, agreement to recontact) and identifying outliers. Interviewers, who were consistently identified as being outside of the expected range, were flagged for remedial action, such as being warned about performance, further discussions with their supervisor, or in extreme cases being removed from the interviewer panel.

5.3 Fieldwork dates and fieldwork management

During 2019-20 the survey was managed on a monthly basis. An equal number of assignments were issued each month (between 136 and 138).

Interviewers were encouraged to start their assignment as early as possible in the month to minimise the time between respondents receiving the advance letter and an interviewer calling. Interviewers had until the end of the calendar month to cover all the addresses in their assignment and report final outcomes.

Once all the issued addresses had been covered and all electronic outcomes returned to the office, a decision was taken about re-issuing non-productive outcomes. As a general rule all non-productive addresses (non-contacts, refusals, broken appointments, etc.) were re-issued unless there was a specific reason not to or it was considered not to be cost effective (e.g. only one or two addresses in an assignment). Once the first re-issue period had been completed a decision was taken about whether to re-issue addresses that were still non-productive for a second or third time.

In total, across the year 9,629 addresses were re-issued on the core sample, which represented 17% of the original sample. Of these, 1,245 addresses were issued for a second time (2% of all addresses). The volume of reissues was less than previous years due to the stoppage of fieldwork as a result of COVID-19 (see section 6.2).

Of all the addresses re-issued, 26% were converted into productive outcomes at some stage. Addresses where the original outcome had been a refusal were less likely to be converted (22%) compared with addresses where the original outcome had been a non-contact (32% converted) or another unproductive outcome (35%). Overall, the impact of the re-issue process was to increase the response rate on the core sample from 58.6% after the initial issue to the final response rate of 64.0%.

As a result of this time lag between addresses being issued and interviews being achieved, it is normal practice on the CSEW for the 12 months of issued sample in any one year to be worked in the field over a 15 month period, with each issued quarter of sample being in the field for up to six months. This means that not all interviews are achieved in the quarter in which they are issued. Approximately 80% of interviews are achieved in the same quarter they are issued, with 20% falling into the next quarter. In 2019-20 the fieldwork was scheduled to run from 1st April 2019 to 30th June 2020. However, in the event fieldwork was halted on the 17th March 2020 due to COVID-19 meaning that the normal fieldwork pattern is slightly different compared with previous surveys.

The questionnaire used in the field is always aligned to the survey year, rather than being aligned to the sample issue period. In 2019-20 all interviews carried out between 1st April 2019 and 17th March 2020 were therefore completed using the 2019-20 questionnaire, irrespective of the time period in which the sample was issued. The advantage of this is that the questionnaire is in line with the way in which the data are reported.

Further details of how the quarterly data outputs relate to the issued and achieved sample can be found in [section 8.2](#).

5.4 Advance letter and leaflet

All selected addresses were sent a letter from the Office for National Statistics in advance of an interviewer calling at the address. This explained a little about the survey, why this particular address had been selected and telling the occupiers that an interviewer from Kantar would be calling in the next few weeks. The letter also provided a telephone number and an email address for people to contact to find out more about the survey, to make an appointment for an interviewer to call, or to opt out of the survey. Over the course of the whole year 2,502 people, representing around 4% of addresses issued, opted out of the survey by contacting either Kantar or ONS.

Included with the advance letter was a leaflet from the Office for National Statistics which provided people with some more details about the survey, including findings from the previous survey. The leaflet also tried to answer some questions that potential respondents might have such as issues relating to confidentiality.

A leaflet was also specifically designed for the 10 to 15 year olds that explained in relatively simple terms what the survey was about. This leaflet was not sent to households in advance and was rather handed out by the interviewer in eligible household, usually after conducting the core survey. Much of the detailed information about the survey was omitted from this leaflet on the basis that the 10 to 15 year olds would also have access to the original household letter and leaflet about the survey.

Examples of the advance letters used can be found in Appendix A and a copy of the leaflets (including the leaflet designed for 10 to 15 year olds) can be found in Appendix B of Volume 2.

In Wales, a bilingual copy of the advance letter and survey leaflet were sent to all selected addresses.

5.5 Respondent website

A website with information about the survey was set up, the style and content of which is updated regularly. Respondents can be directed to this website by the interviewer and the website is referenced in all respondent-facing survey materials.

Information displayed on this website include what the survey is about and what types of questions are asked, survey results, confidentiality and data security, the 10-15 year olds survey and a section on frequently asked questions.

The URL for the website is: <http://www.crimesurvey.co.uk/index.html>

5.6 Electronic Contact Sheet (ECS)

All records about the individual addresses issued to interviewers and details about the calls made to those addresses are stored using the Electronic Contact Sheet.

The Electronic Contact Sheet is crucial to the management of the CSEW, both at the level of the individual assignment and for the management of the survey overall. Information from the ECS is transferred securely to Head Office on a daily basis so that overall progress can be monitored and managed

The primary functions of the ECS are as follows:

- To record the days and times that interviewers call at an address and any additional comments.
- To record the outcome achieved at the address at every visit, including all reissue visits.
- To carry out the various random selection procedures as required and record the details- this includes any dwelling unit selection, household selection, and person selection.
- To carry out the screening and selection process for 10 to 15 year olds and record all outcomes associated with this survey.

- To collect some basic information about the area and the selected address (e.g. type of property, condition of the property, whether it is in a Neighbourhood Watch area, etc.). This information is collected by interviewers based on their own observations and, as such, is subjective. Nevertheless, such information does tend to be highly associated with non-response and is also used by ONS as an area-based disorder measure.

5.7 Incentives

Since 2005, a booklet of six first class stamps has been sent with the advance letter as a 'thank you' to people for taking part in the survey and this was the case for the 2019-20 survey. Between October 2017 and March 2018 an experiment was conducted to test the impact of replacing the stamps with a tote bag for a quarter of the survey sample³⁰.

During the 2015-16 survey, a £5 gift card was offered to 10 to 15 year olds who completed the survey. The use of this incentive has continued since then and was used during the 2019-20 survey.

5.8 Fieldwork procedures and documents for the 10-15s survey

All respondents to the 10-15 year olds survey were selected from households already selected to take part in the core survey. Screening for 10 to 15 year olds was only carried out in households where a successful adult interview was achieved. In most cases screening was conducted only on completion of the adult interview, although in some situations it was carried out before the adult interview had taken place.

Where a 10-15 year old was identified in a household, interviewers had first to obtain the permission of a parent or guardian to interview the child and then the consent of the child themselves before starting the survey. Permission was recorded on the ECS by recording the name of the adult giving consent and their relationship to the selected child. In some cases, the adult interviewed on the main survey respondent was not the parent or guardian of the child (for example, an older sibling or a grandparent). In such situations, interviewers were not able to obtain permission to interview the child from the core respondent but had to make contact with the parent or guardian instead.

Interviewers were provided with a parental information card which gave details of the nature and content of the survey. This was presented to parents or guardians as part of gaining permission to interview the child. An example of this card can be found in Appendix B of Volume 2.

Once parental permission was obtained interviewers were instructed to ensure that the 10-15 year old also gave their consent to participate in the survey and that they understood what the survey would be about. In order to emphasise to 10-15 year olds their right to refuse any particular question, they were given a red and green card to use throughout the interview. If they did not want to answer a question, they could simply show the interviewer the red card and that particular question would be coded as a refusal. This technique was developed primarily with the younger age groups in mind as a way to reassure parents.

5.9 Presence of others during the interview

During the interviewer briefing sessions emphasis was placed on the importance of trying, wherever possible, to conduct the interview in private. This generally helps to make the interview run more smoothly, but it also might encourage some respondents to mention certain incidents or events, which they might be embarrassed or worried of talking about in front of others.

Privacy during the interview is a particular concern for respondents who have experienced domestic violence or sexual assault. Where respondents had experienced such incidents in the last 12 months, interviewers had the option of suspending the victimisation module (simply by skipping over it) if they felt it was

³⁰ See 2018-19 Technical Report for details about this experiment.

inappropriate to continue with the questions because of the presence of others in the room. This procedure meant that the interviewer could complete the rest of the questionnaire, rather than having to abandon the whole interview. During 2019-20, a total of 12 victimisation modules were suspended by interviewers for this reason.

Although it is preferable for the interview to be conducted with no-one else present, there are also some situations where the presence of others might improve the accuracy of the information collected. This is particularly the case for incidents of vehicle crime or property crime, where the respondent may not have been personally present, reported the incident to the police, etc. Additionally, in many cases it is simply not be possible for the interview to be conducted without others present in the room.

5.9.1 Presence of others during the adult screener interview

The key point at which the presence of another person could affect the estimate of victimisation is during the initial set of screener questions. Respondents may be less willing to report incidents when other are present during the interview.

Therefore, at the end of these questions, the interviewer recorded whether anyone else was present or not. Table 5.1 shows whether or not anyone else was present in the room during the initial screener questionnaire, when respondents are giving details about their experiences of crime.

Table 5.1 Presence of others during the screener questionnaire, 2019-20 CSEW

	Core sample
	%
No-one present	71
Child(ren) under 16	7
Spouse/partner	18
Other adult	7
<i>Base: All adult respondents</i>	32,028

In 2019-20, seven out of ten (71%) adult respondents were interviewed with no-one else other than the interviewer being present. Where someone else was present, the people most commonly there were the respondent's spouse or partner (18%).

There was little difference between men and women as to whether they completed the interview with no-one else being present (72% of men and 70% of women). Younger respondents were less likely to complete the interview alone (62% of 16-24 year olds). Asian respondents overall were less likely compared with respondents from other ethnic groups to complete the interview alone (64%), although there was a significant gender difference (70% of Asian men and 58% of Asian women completed the survey with no-one present).

However, any patterns by age or ethnicity will also be influenced by household composition. Table 5.2 shows the information from the previous table with single person households identified separately. Over nine in ten (93%) respondents interviewed in single person households were interviewed with no-one else present. In households containing more than one person, around four in ten respondents (38%) were interviewed with someone else present.

Table 5.2 Presence of others during the screener questionnaire by household size and sample type, 2019-20 CSEW

	Single person household	More than one person household
	%	%
No-one present	93	62
Child(ren) under 16	1	10
Spouse/partner	*	25
Other adult	6	8
<i>Bases: All adult respondents</i>	9,650	24,084

5.9.2 Presence of others during the self-completion and assistance given

For those who did the self-completion, the presence of others during this part of the interview was also recorded. Self-completion should offer a respondent a degree of privacy, even when others are present during the interview. Where this was the case, interviewers were briefed to try and 'arrange' the room whenever possible so that the respondent had a degree of privacy - for example, ensuring that the respondent was sitting with the screen facing a wall or was in a position so that no-one else in the room could read the computer screen. However, given the sensitive nature of the modules it could still be the case that some respondents are less likely to report certain things if other people are present during the interview.

Table 5.3 shows that three quarters of adult respondents (74%) who did the self-completion did so when no-one else was present. Thirteen per cent completed the self-completion with a spouse or partner present and 7% did so when children were present in the room.

Table 5.3 Whether anyone else was present or not during the self-completion, 2019-20 CSEW

Core sample	
	%
No-one else	74
Spouse/partner/girlfriend/boyfriend	13
Child(ren) under 16	7
Other household member (adult)	5
Someone else	3
<i>Base: All adult respondents who did the self-completion (inc. interviewer administered)</i>	27,260

Percentages add up to more than 100% since more than one answer could be coded at this question.

Where anyone else was present, the extent to which they looked at or discussed the questions with the respondent was also recorded. In 13% of cases where someone else was present during the self-completion, they looked at or discussed the questions with the respondent; in 6% of cases someone else looked at or read the self-completion with the respondent, while in 7% of cases the respondent discussed the self-completion with other people.

The amount of help or assistance provided by the interviewer during the self-completion part of the interview was also recorded (Table 5.4). Respondents who accepted the self-completion module rarely needed help.

Table 5.4 Amount of assistance given by interviewers with the self-completion questionnaire, 2019-20 CSEW

Core sample	
	%
All done by respondent	91
Help given with one or two questions	6
Help given with more than one or two questions, but less than half	2
Help given with more than half, but not all	1
Help given with all/nearly all	1
<i>Base: All adult respondents who did the self-completion (exc. Interviewer administered)</i>	25,698

5.9.3 Presence of others during the 10-15 year olds interview

The 10-15 year olds interview was more likely to take place in the presence of others than the adult interview, with a parent or guardian being the most likely person to be present during the screener questionnaire. As might be expected, there was a clear association between the age of the child and the likelihood of a parent or guardian being present. Thus, when interviewing a 10 year old a parent or guardian was present in 89% of interviews compared with 66% of interviews with 15 year olds.

Table 5.5 Presence of others during the screener questionnaire, 2019-20 CSEW, 10-15 year olds sample

	Age of child						Total
	10	11	12	13	14	15	
	%	%	%	%	%	%	%
Parent/guardian	89	84	82	76	71	66	78
Other child from household	13	12	12	10	8	6	10
Other adult from household	2	2	3	3	2	4	3
Other non-household child	2	1	2	2	1	1	2
Other non-household adult	2	1	2	1	1	2	2
No one present	10	13	14	22	25	31	19
Base:	387	389	451	395	375	401	2,398

5.10 Length of interview

Timing stamps were placed throughout both the adult and 10-15 year olds questionnaire to allow timing of individual sections. In a small number of cases, the time stamps were invalid due to technical issues although valid times were available for around 98% of interviews.

The average (mean) core interview length in 2019-20 was 48 minutes (median 44 minutes). Eight in ten (80%) adult interviews took 60 minutes or less to complete, a further 16% took between 60 and 90 minutes, and a small proportion (4%) took over 90 minutes to complete.

The main influences on core interview length were whether the respondent had been a victim of crime and whether they answered the self-completion modules. The average interview length for victims of crime was 65 minutes compared with 43 minutes for non-victims. Respondents who completed the self-completion

modules of the survey took on average 8 minutes compared with those who did not³¹. Non-victims who did not complete the self-completion modules had the shortest interview length (35 minutes on average).

The average length of interview by number of victimisation modules completed is shown in Table 5.6. Not unexpectedly, interview length was strongly related to the number of victimisation modules completed, with those completing four or more modules (4% of victims) having an average interview length of around 101 minutes.

Table 5.6 Average time of interview by number of victimisation modules, 2019-20 CSEW

Number of victimisation modules	Average time (minutes)
Non victims	43
All victims	65
1	59
2	75
3	89
4 or more	101
All adult respondents	48

In 2019-20, the average interview length of the 10-15 year olds survey was 27 minutes. As was the case with the core adult interview, respondents who reported being a victim of crime had a longer interview. The average interview length for non-victims was 23 minutes compared with 39 minutes for those who reported being a victim of crime. It is worth noting that the survey in 2019-20 was significantly longer compared with previous surveys, probably due to the introduction of the self-completion module on cybercrime and online behaviour.

³¹ This figure includes the introduction to the computer and the completion of the practice questions.

6. Response rate and reasons for non-response: core sample

6.1 Core survey response rate and non-response

The full response and non-response breakdown for the 2019-20 core sample is shown in Table 6.1 In 2019-20, 9.2% of issued addresses were identified as not being an eligible residential address (known as deadwood). The most common type of deadwood was empty or vacant residential properties, which accounted for 5% of all issued addresses.

Interviewers made contact with either the selected respondent or a responsible adult at 96% of eligible addresses, meaning a non-contact rate of 4%. There were two types of non-contact. The most common (3.6% of eligible addresses) was where no contact was made with anyone at the address despite repeated calls over a lengthy fieldwork period. The remaining addresses classified as non-contact (0.7% of eligible addresses) were where contact was made with someone at the address, but no contact was made with the person selected for interview.

At eligible addresses, the most common reason for not getting an interview was due to a refusal, which accounted for 19.2% of all eligible addresses. The most common types of refusal were where no information about the household was given, meaning that the person selection could not be carried out (6.5%) and where the person selected for interview refused to take part in the survey (5.4%). Instances where refusals were made directly to Head Office, accounted for 5% of all eligible addresses. Proxy refusals (someone refusing on behalf of the selected respondent) were less common (1.1%).

A further 5.3% of eligible addresses were categorised as unproductive for other reasons including broken appointments, people who were ill or away during the period of the survey and people who had inadequate English to complete the survey.

Overall, 32,028 adult interviews were achieved in 2019-20 representing a response rate of 64.0%. Both the total number of interviews achieved and the final response rate were below target as a result of fieldwork being halted in mid-March due to COVID-19. The next section examines the impact of stopping fieldwork early on survey response.

Table 6.1 Core sample response rate and non-response outcomes, 2019-20 CSEW

	N	% of issued	% of eligible
TOTAL ISSUED ADDRESSES	55,102	100	
Deadwood			
Addresses not traced/accessible	445	0.8	
Not built/does not exist	86	0.2	
Derelict/demolished	176	0.3	
Empty/vacant	2,762	5.0	
Second home/not main residence	726	1.3	
Business/industrial	616	1.1	
Institution	162	0.3	
Other deadwood	113	0.2	
TOTAL DEADWOOD	5,086	9.2	
TOTAL ELIGIBLE ADDRESSES	50,016	90.8	100.0
Non-contact			
No contact made with household	1,808	3.3	3.6
No contact with selected respondent	357	0.6	0.7
Total non-contact	2,165	3.9	4.3
Refusal			
Office refusal	2,502	4.5	5.0
Refused all information	3,241	5.9	6.5
Personal refusal	2,681	4.9	5.4
Proxy refusal	564	1.0	1.1
Contact made, no specific appointment	604	1.1	1.2
Total refusal	9,592	17.4	19.2
Other unproductive			
Broken appointment	1,033	1.9	2.1
Temporarily ill/incapacitated	367	0.7	0.7
Physically or mentally unable	375	0.7	0.7
Away/in hospital	334	0.6	0.7
Inadequate English	340	0.6	0.7
Other unsuccessful	188	0.3	0.4
Total other unsuccessful	2,637	4.8	5.3
Fieldwork not completed			
Address not started in field	562	1.0	1.1
Address started – unknown eligibility	2,140	3.9	4.3
Address started – know eligibility	892	1.6	1.8
Total fieldwork not completed	3,594	6.5	7.2
TOTAL UNPRODUCTIVE	17,988	32.6	36.0
Full interviews	32,010	58.1	64.0
Partial interviews ³²	18	0.0	0.0
TOTAL INTERVIEWS	32,028	58.1	64.0

³² Interviews completed to the end of the demographics module (not including the self-completion modules) were coded as full interviews. Interviews stopped before the end of the demographics module were coded as partial interviews. Only cases meeting the definition of usable partial were included on the dataset.

6.2 Impact of COVID-19 on fieldwork and survey response rates

Face to face fieldwork on the CSEW was halted on 17th March 2020 in response to the COVID19 outbreak. This was in line with the complete cessation of in-home interviewing across the whole research industry. As a result of this the issued sample for the CSEW could not be worked to its usual conclusion. In some cases, issued addresses had interim outcomes, indicating that the interviewer had visited an address but had not yet secured a final outcome and so intended to make further visits; in other cases an interviewer had not yet visited an address and so no outcome had been recorded. The table below shows the status of the issued sample in the field at the point fieldwork was stopped.

Table 6.2 Status of sample in the field at the point of fieldwork cessation

Issue period	October – December 2019	January- March 2019
Fieldwork period	1 st October 2019 – 31 st March 2020	1 st January 2020 – 30 th June 2020
Fieldwork halted	17 th March 2020	17 th March 2020
Issued addresses	14,001	13,542
Addresses with final original outcome	13,997	9,952
Addresses with interim outcome only	4	3,028
Addresses with no outcome	0	562
Proportion of addresses reissued	19%	0%

In addition to this, there were addresses which already had a final unproductive outcome on the original sample issue, but where the plan would have been to reissue this to another interviewer. On average approximately 25% of unproductive addresses at original issue are reissued on the CSEW each quarter. Due to the cessation of fieldwork the normal reissuing pattern could not be completed and so a far lower proportion of addresses were reissued than normal, including no addresses being reissued from the January-March 2020 issued sample.

In calculating the response rates for the 2019-20 survey both an unadjusted and adjusted response rate were examined:

- The unadjusted response rate assumes that all unproductive addresses with unknown eligibility are eligible and so are included in the response rate calculation. This is the approach which has been used for calculating response on previous surveys. Since the proportion of addresses with unknown eligibility is normally fairly small by the end of fieldwork this has a negligible impact on the final response rate.
- The adjusted response rate applies an estimate of eligibility to all addresses where eligibility is unknown. This is calculated based on the eligibility rate of all sample with a known final outcome. This latter approach is the one recommended by AAPOR for calculating response rates³³.

Table 6.3 shows the final response rates achieved on the 2019-20 survey. This shows that while the overall annual response rate is lower compared with the previous year by about five percentage points, adjusting

³³ https://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf

the response rate to account for eligibility made little difference to the final response – only increasing it from 64.0% to 64.4%. However, the quarterly response rates show the impact of stopping fieldwork early. While the first two quarters of 2019-20 achieved a response rate of just over 70%, the third quarter had a slightly lower response of 68%, and the final quarter of the year had a response of 49%. This reflects the fact that while fieldwork on the third quarter of the year was only slightly affected by the sudden stoppage, fieldwork in the last quarter was greatly curtailed.

Table 6.3 Adjusted and unadjusted response rates by quarter of issue, 2019-20

	2018-19	2019-20				Annual
		Q1	Q2	Q3	Q4	
Unadjusted response	69.4%	70.5%	70.3%	67.9%	47.7%	64.0%
Adjusted response	n/a	70.5%	70.3%	68.0%	48.6%	64.4%

6.3 Response rate and reasons for non-response: 10-15 year olds sample

Table 6.4 shows the screening and response outcomes for the 10-15 year olds sample. During 2019-20, interviewers were required to screen for 10 to 15 year olds at all of their core sampled addresses where a core interview was conducted.

After accounting for deadwood addresses, 36.0% of addresses which were issued for the core survey were not screened for 10-15 year olds because the outcome at the core address was an unsuccessful outcome. Interviewers identified at least one 10-15 year old at 7.8% of addresses where screening was successfully carried out. Among those households where an eligible respondent was identified the response rate achieved was 57.3%.

The level of non-contact (3.7%) was broadly similar to the level achieved on the core sample but the level of refusals was higher at 34.8%.

The response rate achieved on the 10 to 15 year olds survey does not take into account households where it was not known whether a 10-15 year old was present because of non-response to the core sample. As such while the total number of interviews achieved was lower than the target due to the early cessation of fieldwork the response rate is unaffected by this. However, when the non-response to the core survey is taken into account the true estimated response rate for the 10-15 year olds survey is 37%. Even allowing for the cessation of fieldwork this is a lower response than previous surveys and was largely driven by an increase in refusals, especially parental refusal.

Table 6.4 Response rate and non-response outcomes 10-15 year olds survey, 2019-20 CSEW

	N	% of issued eligible addresses	% of screened households	% of eligible households
TOTAL ADDRESSES FOR SCREENING	55,10	100		
<i>Core deadwood addresses</i>	5,086			
TOTAL ELIGIBLE ADDRESSES FOR	50,01	100		
No screening attempted (eligibility unknown)	18,00	36.0		
Screening information refused (eligibility unknown)	0	0.0		
Total unknown eligibility	18,00	36.0		
Total households screened for 10-15 year olds	32,00	64.0	100	
Screened households with no 10-15 year old	28,08	56.1	87.7	
Screened households with a 10-15 year old	3,926	7.8	12.3	
Total screened households with a 10-15 year	3,926		100	
No contact with selected respondent	105			2.7
No contact with parent/guardian	40			1.0
Total non-contact	145			3.7
Office refusal	4			0.1
Parent/guardian permission refusal	854			21.8
Personal refusal	344			8.8
Proxy refusal	75			1.9
Contact made, no specific appointment	90			2.3
Total refusal	1,367			34.8
Broken appointment	42			1.1
Temporarily ill/incapacitated	4			0.1
Physically or mentally unable	67			1.7
Away/in hospital	21			0.5
Inadequate English	5			0.1
Other unsuccessful	25			0.6
Total other unsuccessful	164			4.2
TOTAL UNPRODUCTIVE	1,676	3.4		42.7
Full interviews	2,249			57.3
Partial interviews	1			0.0
TOTAL INTERVIEWS	2,250			57.3

6.3.1 Core response rates by Government Office Region

Table 6.5 shows the different response rates and reasons for non-response achieved by region in 2019-20. This shows that across regions the response rate ranged from 74% in the North East to 59% in the East of England.

Table 6.5 Core sample response rates and non-response by Government Office Region, 2019-20 CSEW

	North East	North West	Yorkshire & The Humber	East Midlands	West Midlands	East of England	London	South East	South West	Wales
Percentage of eligible addresses (%):										
Non-contact	4.3	3.5	4.1	4.1	5.6	3.8	7.3	3.2	3.4	3.9
Refusal	13.6	18.0	19.5	20.2	18.5	21.9	17.5	20.7	21.8	16.1
Other unproductive	3.9	4.2	4.9	5.9	6.1	6.6	6.0	4.0	5.6	4.9
Achieved interview	73.5	66.8	66.5	64.2	59.7	59.2	61.1	66.0	61.2	68.5

6.3.2 Core response rate by Police Force Area

As outlined in [section 1.1](#) the aim was to achieve a minimum of 625 interviews in each PFA, with larger sample sizes in the most populous areas. In order to achieve this sample size within each PFA the amount of sample issued was based on actual average deadwood rates and response rates over the previous year.

[Table 6.6](#) below shows the actual number of interviews achieved in each PFA and the response rates. Due to the early cessation of fieldwork the total number of interviews achieved in each PFA was slightly lower than target, but the shortfall was fairly evenly spread across all areas.

Table 6.6 Core sample achieved interviews and response rates by PFA, 2019-20 CSEW

PFA	Target	Achieved	Response rate
	N	N	%
Avon & Somerset	850	799	65.0
Bedfordshire	625	525	57.3
Cambridgeshire	625	539	60.1
Cheshire	625	619	67.6
Cleveland	625	599	70.1
Cumbria	625	587	68.4
Derbyshire	625	587	66.0
Devon & Cornwall	943	845	61.6
Dorset	625	576	62.3
Durham	625	613	78.0
Dyfed Powys	625	581	71.6
Essex	903	849	57.7
Gloucestershire	625	509	53.7
Greater Manchester	1,410	1,303	67.1
Gwent	625	573	71.0
Hampshire	992	924	71.9
Hertfordshire	625	580	61.2
Humberside	625	579	67.6
Kent	893	801	61.5
Lancashire	774	717	67.8
Leicestershire	625	548	56.7
Lincolnshire	625	613	70.9
Merseyside	903	847	64.2
Metropolitan and City of London	3,899	3,659	61.1
Norfolk	625	554	62.7
North Wales	625	626	69.6
North Yorkshire	625	609	64.3
Northamptonshire	625	545	60.4
Northumbria	781	745	73.0
Nottinghamshire	625	602	67.9
South Wales	678	603	62.8
South Yorkshire	708	651	62.5
Staffordshire	625	573	67.5
Suffolk	625	536	57.5
Surrey	625	647	68.8
Sussex	853	911	66.9
Thames Valley	1,146	1,045	62.8
Warwickshire	625	509	56.3
West Mercia	625	559	60.2
West Midlands	1,355	1,247	57.9
West Yorkshire	1,162	1,117	69.9
Wiltshire	625	577	61.8

6.3.3 Core response rates by type of area

Since large administrative areas such as regions contain a variety of different types of area it is useful to examine response to the survey broken down by area type. [Table 6.7](#) shows the response rates and reasons for non-response by different types of area, showing that overall response rates tended to be lower in areas categorised as inner city compared with non-inner city areas (59% and 65% respectively). This difference in response rate explains why the current CSEW data includes a weight to correct for differential response rates between those areas defined as inner city and non-inner city ([see section 9](#)).

Similarly, the response rate in urban areas was slightly lower compared with that achieved in rural areas (63% and 67% respectively). Response also varied significantly by ACORN³⁴ Category, being highest in areas classified as 'Affluent achievers' (68%) and lowest in areas classified as 'Urban adversity' (60%). There was similar variation in response by Output Area Classification, ranging from 69% in 'Rural residents' to 55% in 'Cosmopolitans'. Looking at the differences in response rates by types of area shows how most of the response differential is due to variation in the non-contact rate, while the refusal rate tends to be fairly consistent. Thus, while the refusal rate varied between 17% and 21% in the different types of areas shown in [Table 6.7](#), the non-contact rate varied from 2% to 12%.

³⁴ For details of ACORN categories please see: <http://acorn.caci.co.uk/downloads/Acorn-User-guide.pdf>

Table 6.7 Core sample response rates and non-response by types of area, 2019-20 CSEW³⁵

	Non-contact	Refusal	Other unproductive	Achieved interviews
Percentage of eligible addresses				
	%	%	%	%
Inner city ¹	7.6	17.6	6.6	59.0
Non-inner city	4.0	19.3	5.1	64.6
Urban ²	4.7	19.1	5.6	63.4
Rural	3.0	19.3	4.1	66.6
ACORN Category				
Affluent achievers	2.5	20.0	3.5	68.2
Rising prosperity	8.6	17.8	5.1	59.9
Comfortably communities	2.9	19.9	4.8	66.0
Financially stretched	4.9	18.6	6.4	62.5
Urban adversity	6.2	18.3	7.1	59.6
Output Area Classification				
Rural residents	2.2	19.3	3.6	68.5
Cosmopolitans	12.2	17.2	5.1	54.9
Ethnicity central	9.3	17.0	7.2	56.1
Multicultural metropolitans	5.5	17.0	7.1	62.7
Urbanites	4.3	20.3	5.1	63.5
Suburbanites	2.1	20.6	3.7	67.5
Constrained city dwellers	6.0	18.5	7.6	60.2
Hard pressed living	3.7	19.1	5.7	64.4

¹ Inner city is based on the CSEW definition that has been used for many years. See [section 9](#) for more details.

² This is based on the ONS definition of urban-rural areas, where urban is classed as 'urban –sparse' and 'urban –less sparse' and all other areas are classed as rural

³⁵ Due to suspension of fieldwork, a small proportion of addresses did not receive a final outcome. As a result, percentages do not add up to 100.

6.4 Response to the self-completion questionnaire

The last part of the core adult questionnaire involved a self-completion module which was asked of all respondents aged 16-74. In 2019-20 there were four self-completion modules on the survey:

- Use of illicit drugs and drinking behaviour
- Gangs and personal security (Groups A and B aged 16-29 years old)
- Experience of domestic violence, sexual victimisation, and stalking
- Nature of serious sexual assault

Although respondents were encouraged to use the computer themselves, if they did not want to use it for any reason, interviewers were allowed to administer the modules provided that no-one else was present in the room. Where the self-completion part of the survey was administered by the interviewer the domestic violence, sexual victimisation and stalking modules were not completed, since these questions were considered too sensitive to be read out by the interviewer.

Table 6.8 shows that 94% of eligible respondents answered the self-completion module, with 89% of them entering their answers directly into the laptop themselves and 5% asking the interviewer to enter their answers for them.

Table 6.8 Response to the self-completion module, 2019-20

Core sample	
	%
Refused	5.6
Completed by interviewer	5.4
Accepted by respondent	89.0
Overall self-completion response	94.4
<i>Base</i>	28,902

Table 6.9 shows how response to the self-completion questionnaire varied according to the demographic characteristics of respondents.

There was no difference between men and women in terms of response to the self-completion. Older respondents were more likely than younger ones to ask the interviewer to enter their answers for them (10.8% of respondents aged 65-74 compared with 2.5% of 16-24 year olds).

Some of the most noticeable differences were between respondents from different ethnic groups. Only 5.2% of White respondents refused to do the self-completion compared with 8.6% of Asian respondents and 10.4% of respondents who belong to an 'other ethnic group'. Black, Asian and 'other ethnicity' respondents were more likely than White respondents to ask the interviewer to enter their answers for them.

There were also some differences by socio-economic classification, with respondents from routine and manual occupations being slightly less likely than those from managerial and professional occupations to answer the self-completion (93.0% compared with 96.5%). Refusal rates were highest for respondents who have never worked or are long-term unemployed (17.4%). Respondents who have never worked or are long-term unemployed were also more likely than those from managerial and professional occupations to ask the interviewer to enter their answers for them (14.6% and 2.3% respectively).

Table 6.9 Response to the self-completion questionnaire by socio-demographic characteristics of respondents (core sample), 2019-20 CSEW

	Refused	Completed by interviewer	Accepted by respondent ¹	Overall self-completion response	Bases: N
	%	%	%	%	
Sex					
Male	5.7	5.9	88.4	94.3	13,350
Female	5.6	5.0	89.5	94.4	15,552
Age					
16-24	4.2	2.5	93.3	95.8	2,103
25-44	5.6	3.0	91.4	94.4	10,287
45-64	5.3	5.6	89.1	94.7	11,139
65-74	6.8	10.8	82.5	93.2	5,373
Ethnicity					
White	5.2	5.2	89.7	94.8	25,266
Mixed	6.4	5.0	88.6	93.6	404
Asian	8.6	8.2	83.2	91.4	1,971
Black	7.1	5.9	87.1	92.9	904
Other ethnic group	10.4	7.8	81.7	89.6	268
NS-SEC					
Higher managerial, administrative & professional	3.5	2.3	94.2	96.5	10,902
Intermediate occupations	5.1	5.1	89.8	94.9	6,478
Routine & manual	7.0	8.5	84.4	93.0	9,269
Never worked and long-term unemployed	17.4	14.6	68.0	82.6	958

¹ Respondent used the laptop on their own

Table 6.10 shows the reasons given by respondents either for refusing the self-completion module or for asking the interviewer to enter their answers for them.

Running out of time was the most common reason cited for respondents refusing to complete the self-completion (mentioned by 41%). A dislike of computers was the most common reason why respondents asked the interviewer to enter their answers for them (mentioned by 43%).

Table 6.10 Reasons for refusing self-completion questionnaire or for completion by interviewer (core sample), 2019-20 CSEW

	Refused	Completed by interviewer	Total
	%	%	%
Don't like computers	14.2	43.3	28.5
Ran out of time	40.8	10.9	25.9
Couldn't be bothered	3.8	3.8	3.8
Language problems	12.3	10.9	11.6
Children in room	8.8	3.6	6.2
Disability	3.6	9.7	6.5
Eyesight problems	2.7	11.7	7.1
Respondent unwell	7.4	10.3	8.9
Interview already too long	28.9	8.6	18.8
Could not read/write	2.1	4.6	3.4
Confidentiality worries	7.6	3.0	5.4
Other people in room	3.5	1.1	2.3
Objected to study	2.0	0.3	1.2
Other reasons	13.3	8.1	10.7
<i>Bases:</i>	1,622	1,562	3,212

Percentages add up to more than 100% since more than one answer could be coded at this question

7. Data processing

7.1 Offence coding

The CSEW Offence Coding System, which was originally developed in 1982 as part of the first Crime Survey, is designed to replicate as far as possible how incidents are classified by the police. The survey counts crime according to the victim's account of events, rather than requiring criminal intent to be proven. This is reflected in how the police record crimes under the National Crime Recording Standard using the Counting Rules³⁶. It should be noted, however, that the Counting Rules evolve and change over time, and while efforts are made to reflect these changes in the survey, there are always likely to be some discrepancies between the two systems.

In order to classify offences, detailed information is collected about the incidents reported by respondents in the victimisation modules. Once the data is returned to the office, all victimisation modules are reviewed by specially trained coders to determine whether what has been reported represents a crime or not and, if so, what offence code should be assigned to the crime.

Apart from some minor changes, the code frame and the instructions to coders on the core survey (see Volume 2 for a copy of the Coding Manual) have remained largely unchanged since 1982. The current operational procedures used for assigning codes have been in place since 2001. In 2010 the coding process was updated to include the coding of offences against 10 to 15 year olds, while in 2015 it was updated to include the classification of fraud and cyber offences. Neither of these changes affected the way in which non-fraud incidents affecting adults were coded.

The coding manual itself is reviewed annually. Most updates are minor modifications to account for new scenarios that evolve over time and to reflect changes in the Counting Rules. However, in October 2018, a more significant update was incorporated to change the classification of offences related to identity theft. Prior to the change these incidents were recorded as computer misuse offences as a result of unauthorised access to the victim's personal details. After the change was applied these offences were recorded as 'other fraud' offences, reflecting the fraudulent use of a victim's details to apply for a loan or another type of credit agreement.

The current Offence Coding System consists of the following steps:

1. For each victimisation module a summary (called an RTF) is produced drawing together the key information from the module into a single easy reference document. This allows the coders to review each incident as a whole and make a judgement on the most appropriate code to allocate based on the totality of the information.
2. In addition to these summaries, the coder use a specially developed computer assisted questionnaire to help them arrive at a final offence code for each incident.
3. In addition to recording an offence code for all fraud crimes, coders record whether the offence meets the criteria for being a cybercrime or not.

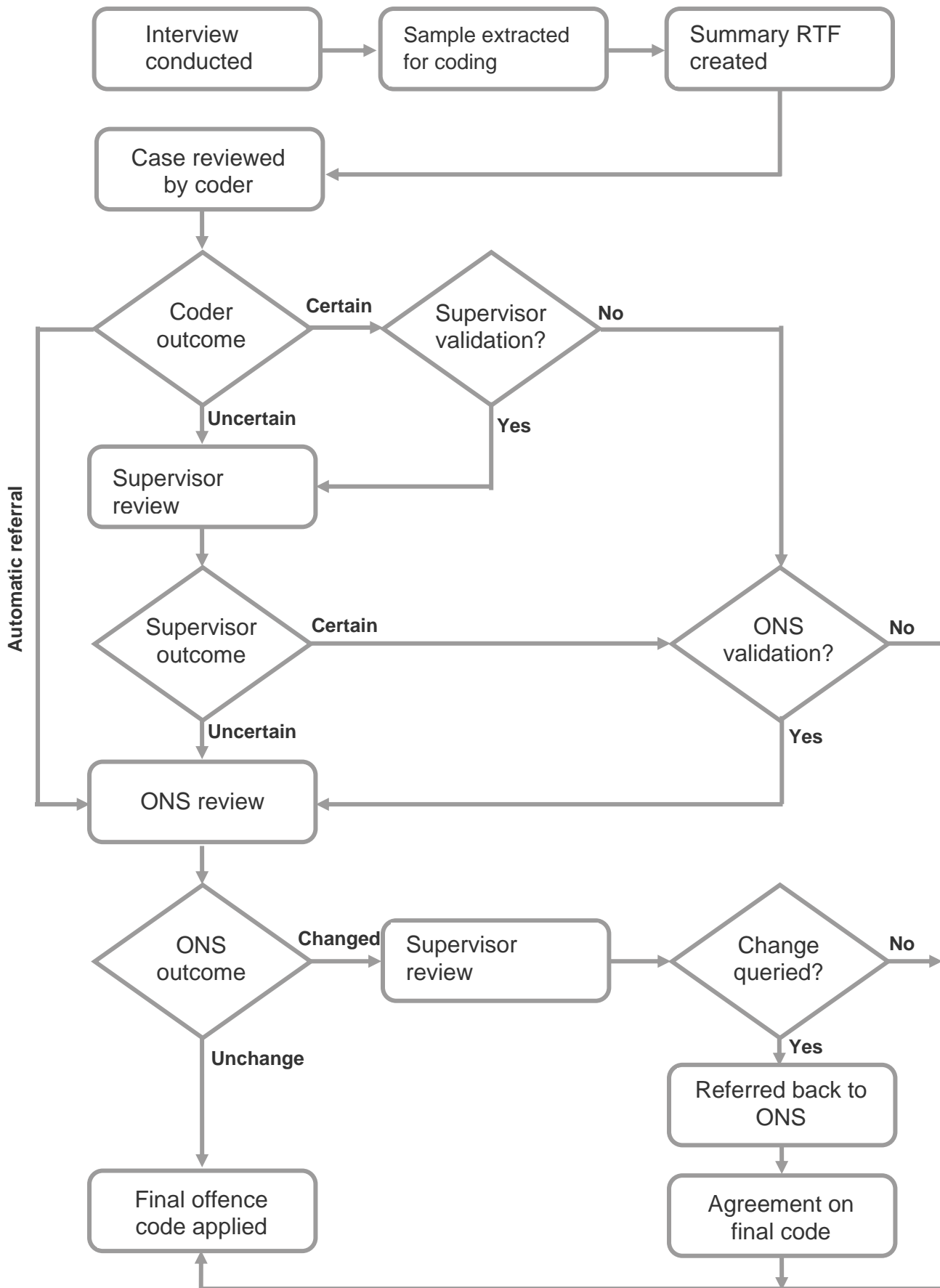
³⁶ https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/877783/count-general-apr-2020.pdf

4. A supervisor checks any codes that the original coder is uncertain about. Additionally, 5% of codes where the coder is certain of the outcome are also checked by a supervisor as a further quality check. These are systematically selected from all cases that have been coded (i.e. every *n*th case) in a particular period.
5. A further quality check is carried out by a team at the Office for National Statistics who examine:
 - Any codes where Kantar is uncertain about.
 - Certain types of incident that are automatically referred (e.g. arson).
 - A proportion (minimum of 5%) of certain codes, as part of a general quality control check. Again, these cases are systematically selected from all cases that have been coded.

The result of this process is that every victimisation module has a final offence code assigned to it. Although the coding rules are broadly similar, separate instructions exist for the coding of traditional (non-fraud) incidents, fraud and computer misuse incidents, as well as incidents affecting 10-15 year olds.

A flow chart of the Offence Coding process is shown in Figure 7.1 and the offence coding system is explained in more detail below.

Figure 7.1 CSEW Offence Coding Flowchart



7.2 The offence coding task

Coders are provided with a summary sheet (called an RTF) of the key variables from each victimisation module and this information forms the basis of the coding. This summary sheet includes the open-ended description collected during the interview, as well as some of the key pre-coded questions in the survey which feed into the classifying of offences. It is important that the coders can consider all the information in its totality because sometimes the information collected may not be entirely clear or some of the information may appear contradictory or inconsistent. While a lot of emphasis is placed on the training and briefing of interviewers about collecting comprehensive and accurate data, inevitably there are cases where coders must make judgements about which bits of information to give priority to.

To assist with their task, coders use a specially designed computer assisted questionnaire to carry out the coding. This questionnaire consists of several different modules each of which relate to a high-level offence category (assault, burglary, theft, criminal damage, fraud, etc.). For each case coders must select an offence module to start with. Once in a module the questionnaire programme asks the coders a series of questions about the incident, and they are able to use the information from the RTF to record an answer. The questionnaire is structured like a flow chart to take account of the major rules that apply to offence coding (such as the priority of codes). By answering the sequence of questions based on the information provided in the victimisation module, the coder either reaches an offence code or is directed to another module to repeat the process.

The coders are also provided with a coding manual. The manual contains all the rules that govern offence coding plus further guidance by using specific examples. The manual also provides flow-charts that show how the coding questionnaire works, so that coders can see how they reached a particular offence code on the basis of the answers that they input. The coding manual is kept updated both in terms of major changes to the survey (such as the incorporation of coding guidelines for the 10 to 15 year olds survey in 2010 and the incorporation of fraud and cybercrimes in 2015, as well as being updated to add additional detail and guidance based on the experience of the coders and other feedback.

The current Offence Coding Manual can be found in Appendix I in Volume 2 of the 2019-20 Technical Report.

Once a coder arrives at an offence code using all of the resources outlined above, they also record whether they are certain or uncertain that it is the right code. Any case where the coder is uncertain is automatically referred to a supervisor for checking. In addition, supervisors check a minimum of 5% of codes which coders are certain about as part of the quality assurance process.

7.3 Quality assurance by ONS coders

All cases where coders are uncertain about the correct code to assign are automatically referred to ONS. In addition to this, a minimum of 5% of all codes which coders are certain about are selected to be sent to ONS for quality control checking. These are selected in a systematic fashion by selecting every *n*th case in each two-week period.

All quality assurance checks carried out by researchers at ONS take place through an online offence coding portal. Victimisation modules to be checked by ONS staff are uploaded to the portal every week. The offence coding portal contains the unique serial number of each victimisation module, the code that the coder (and supervisor if applicable) has given the incident, how certain the coder (and supervisor) is about the code, and any notes that the coder has added about why they are uncertain. The RTF summary document providing the key variables from the victimisation module are also available from the portal for ONS staff.

Researchers at ONS review each of the victimisation modules sent to them via the portal and add any comments they have on each case. For all cases they either accept the code given by the coder or suggest

a different code. These codes then appear on the offence coding portal so that the coders can see the changes that have been made. Apart from making the process more efficient the portal also ensures a complete audit trail for every case.

Once all cases have been reviewed by ONS staff the coding team at Kantar review all cases where a code has been changed. Particular attention is paid to cases where ONS has changed a code that Kantar coders had marked as “certain”. If the Kantar coders disagree with the ONS coding decision, it is flagged up in the coding portal to both Kantar researchers and ONS researchers for further consideration and discussion. This approach of iterative review is continued until everyone is agreed on the final outcome code.

As part of the 2019-20 survey, a total of 1,847 cases were sent to ONS for checking, which represented about 17% of all adult victimisation modules (both traditional and fraud cases). Overall 1,249 traditional (non-fraud) cases were sent for checking (18% of all cases) and 598 fraud cases were sent (17% of all cases).

7.3.1 Traditional (non-fraud) cases referred to ONS

Of the 1,249 traditional (non-fraud) modules sent to ONS:

- 123 cases were automatically referred. This covers cases of aggravated burglary, duplicate cases and cases where the victimisation module was invalid;
- 120 cases were sent because the Kantar coders were uncertain about the code; all uncertain codes are automatically referred;
- 539 cases were sent as part of the systematic quality control check; and
- 467 cases were related victimisation modules. To ensure that those checking offence codes have complete information, all the victimisation modules related to an individual respondent are sent to ONS, rather than just the single module under consideration.

Of the 1,249 non-fraud modules referred to ONS, 52 cases initially had their code changed by ONS, representing 4% of all cases sent. This level of change has been fairly static across survey years suggesting a degree of stability in the offence coding process.

In all cases where ONS staff changed a code that Kantar coders or supervisors had been certain about, the change was reviewed by a coding supervisor and if there was still disagreement over the final code it was referred back to ONS for further review based on providing additional information on the reasons for reaching a particular code. At the end of this iterative process, only 10 codes were changed from the code originally allocated by the coder or supervisor. In one further case, neither the Kantar or ONS code was deemed to be correct and a new code was applied.

7.3.2 Fraud cases referred to ONS

Of the 598 fraud cases sent to ONS for checking as part of the 2019-20 survey:

- 40 cases were automatically referred to ONS. This covers duplicate cases and cases where the victimisation module was invalid;
- 129 cases were where the Kantar coders were uncertain about the code; all uncertain codes are automatically referred;
- 307 cases were sent as part of the systematic quality control check; and
- 122 cases were related victimisation modules.

Of the 598 fraud modules sent to ONS, 35 cases initially had their code changed by ONS staff, representing 6% of all cases sent. However, following further review and discussion only one case was changed from the original code, while a further two cases had a new code applied to them.

7.4 Final Offence Code

The SPSS data sets delivered to ONS on a quarterly basis include all the offence codes that have been given to each victimisation module at each stage of the coding process. This ensures an audit trail exists for each case. The final offence code is derived using a priority ordering system, whereby the ONS code takes priority over the supervisor code, which takes priority over the original code assigned by the coder. The variables on the data file are:

VOFFENCE	Code assigned by the original coder
SOFFENCE	Code assigned by the supervisor (if coded)
FINLOFFC	Code assigned by the ONS team (if coded)
OFFENCE	Final offence code

7.5 Checks on final offence code

Once the SPSS data sets are run some further consistency checks are applied to the final offence codes, checking the offence codes against key pre-coded variables in the victimisation module. The purpose of this is to highlight cases where some of the pre-coded data seems potentially anomalous with the final offence code. Such anomalies can arise because occasionally the information reported by the respondent is not consistent, or even seems contradictory. In particular, there can be inconsistencies between the verbatim description of the incident and subsequent pre-coded questions. While interviewers are carefully briefed to try and be aware of such inconsistencies arising during the interview it is inevitable that some will be missed. Consistency checks within the actual questionnaire script to try and pick up anomalies are not possible when a verbatim description is involved.

The consistency checks carried out are as follows:

- Assaults where no force or violence is recorded as having been used
- Burglary where entry to the property is recorded as being authorised
- Car thefts where no car is recorded as being stolen, or where the police were not informed
- Sexual assaults where there is no sexual element to the assault recorded
- Snatch thefts where the item stolen is not recorded as being held or carried
- Other thefts where the item stolen is recorded as being held or carried
- Wounding where no injury is recorded as being sustained
- In scope offences where the offender is perceived by the victim to be mentally ill
- Thefts where nothing is recorded as having been stolen
- Vandalism where no damage is recorded
- Threats where no threat is recorded

Further checks were added in 2015-16 to check the consistency of the fraud coding:

- Computer virus reported where the offence is not classified as a computer virus
- Computer virus where no virus is reported
- Unauthorised access to personal information with loss of money reported
- Fraud with no loss but a loss has been reported
- Checks that the respondent has been correctly identified as a specific intended victim
- Cyber flag checks where inconsistent reporting is evident:
 - Computer virus but no cyber element is reported
 - Classified as a cybercrime but no cyber element is reported
 - Not classified as a cybercrime but a cyber element is reported.

All cases that fail these checks are examined individually by a researcher and, if changes are required the revised code is reviewed by a coding supervisor. Where clear anomalies in the data do exist, it is up to the judgment of the researchers to decide which bits of information should be prioritised in arriving at the final agreed offence code. In such cases, greater credence tends to be given to a good verbatim description of the incident over the answers to specific pre-coded questions where for example anomalies may be a result of interviewer mis-keying or respondent misreporting.

Experience of running these checks shows that most flagged cases do have the correct offence codes, but a few may be amended each quarter as a result of this additional check.

7.6 Variability test

In addition to the verification measures outlined above regular coder variability tests are undertaken by the entire coding team across Kantar and ONS every three to four years. The latest test was conducted in 2017 and is reported in the 2017-18 CSEW Technical Report.

7.7 Other coding

In addition to the offence coding, coders also looked at all questions where an “other –specify” had been given as an answer. The aim of this exercise, commonly known as back coding, is to see whether the answer given can actually be coded into one of the original pre-coded response options. This coding is done in Ascribe, a Windows based coding package.

Coders are provided with the code frames used in the questionnaire as a starting point for coding each year. Since most of the questions have been used in previous years of the survey, the code frames are already well developed and there is little need to add new codes to the frames. However, if the coding supervisor feels an extra code is needed, this is flagged up to researchers who approved any changes before they are implemented.

As, with the offence coding a minimum of 5% of all cases are checked by supervisors as part of the standard quality assurance process.

7.8 Coding of occupation and socio-economic classification

Occupation details are collected for all respondents, either relating to their current job or to their last job if the respondent is not currently employed but had worked at some time in the past. Occupational details of the Household Reference Person are also collected where this is not the same person as the respondent.

For 2019-20 occupations were coded to the Standard Occupational Classification 2010 (SOC2010). All occupational coding is done centrally by specialist coders once the data is returned by interviewers. Coding is done using CASCOT, a package widely used to code occupation, with coders using the ONS manuals for reference.

As well as occupation codes, National Statistics Socio-Economic Classification (NS-SEC) is added to the file for all respondents and Household Reference Persons. NS-SEC categories are derived automatically using an algorithm which was developed from the documentation provided by the Office for National Statistics. Both the NS-SEC operational categories and the NS-SEC analytical categories are derived.

Details of the NS-SEC categories can be found in Appendix J of Volume 2.

7.9 Data processing on the 10 to 15 survey

The offence coding system used on the 10 to 15 year olds survey is based on the system designed for the core survey, but is adapted to be suitable for the types of incidents experienced by 10 to 15 year olds. The main difference is the inclusion of two different counts of crime to reflect severity of crime. The **‘preferred**

measure’ takes into account factors affecting the severity of an incident, such as relationship between offender and victim, level of injury, value of items stolen or damaged. Incidents judged not to be severe are excluded from this measure. The **‘broad measure’** includes all minor offences, including those between children and family members or ‘playground incidents’, which would not normally be treated as criminal matters.

7.10 Offence coding system on the 10 to 15 year olds survey

The process for coding offences on the 10 to 15 year-old survey is exactly the same as that used on the adult survey and the same coding portal is used. The Offence Coding Manual used by coders contains some additional guidance on how to code incidents on the 10 to 15 year-old survey, including the main differences to look out for when coding.

As with the adult survey all cases which the coders are uncertain about are referred to ONS for further verification. In addition, 10% of all codes which Kantar coders are certain about are systematically selected and sent to ONS for quality control checking. This is a higher proportion of cases than is sent on the adult survey, which partly reflects the fact that the offence coding system has been developed more recently compared with the adult survey and partly reflects the lower volume of coding, which means coders do not build up the same level of experience so quickly.

In total, 141 cases were sent to ONS for checking as part of the 2019-20 10 to 15 year olds survey.

Of the victimisation modules sent to ONS:

- 28 cases were automatically referred to ONS. This covers cases including any sexual element, duplicate cases and cases where the victimisation module was invalid;
- 12 cases where the Kantar coders were uncertain about the code;
- 53 cases were part of the quality control check; and
- 56 were related victimisation modules

Of the 141 modules referred to ONS, two had their codes changed by ONS, representing 1% of all cases sent. After further review these changes were accepted and a further case had a new code applied after discussion.

7.11 Final offence code

The SPSS set delivered to ONS includes all the offence codes that have been given to each victimisation module at every stage of the coding process. It also includes an additional variable ‘Offclass’ which defines whether an incident is classified as a ‘relatively minor’ incident or as a ‘relatively serious’ incident. This classification is not part of the coding process but is derived in SPSS based on answers to a small set of questions coded by the coders covering:

- Whether there was intention to steal, hurt or damage
- Whether the victim knew the offender
- The level of any hurt inflicted or cost of items stole or damaged³⁷

An additional variable Offclass2 is included in the dataset (added in 2013-14) which classifies the offence as a ‘relatively minor’ incident or as a ‘relatively serious’ incident based on the responses to questions about intent added to the questionnaire in April 2012 as well as the coded answers given.

³⁷ The guidelines for defining the level of hurt inflicted or cost of any damage or theft are included in the coding manual in Volume II of the 2011/12 Technical Report (Appendix H, pages 9 and 10).

The same consistency checks that are run on the adult SPSS data are run on the 10 to 15 year olds SPSS data to check the offence code against the rest of the data.

8. Data Output

8.1 Introduction

The main outputs provided to ONS are SPSS data files that are delivered on a quarterly basis. Separate data files are provided for the core sample and the 10 to 15 survey sample. For each type of sample, two data files are provided: The Non-Victim File and the Victim File.

The **Non-Victim File (NVF)** is produced at the level of the individual respondent and contains all questionnaire data and associated variables, except for information that is collected in the victimisation modules. Data for both victims and non-victims are included on the Non Victim File.

The **Victim File (VF)** is produced at the level of the individual incident and contains all the data collected in the victimisation modules. Thus, an individual respondent who reported three crimes and completed three victimisation modules would have three separate records in the Victim File. All generated victimisation modules were included on the file, including cases where the module either had been suspended or where the reference period was out of scope. Although such records contain no information and are not used for analysis, it is useful to keep these on the file to monitor the number of modules that fall into these categories.

8.2 Delivery of data output

During 2019-20 survey, four data files were supplied to ONS on a quarterly basis (April 2019 to March 2020). Data was supplied on a 12-month rolling basis, meaning that each new data delivery was updated by adding the newest quarter of data and deleting the oldest quarter of data.

In addition to the achieved sample, a data file of the entire 2019-20 issued sample was supplied to ONS alongside the annual April 2019-March 2020 data file. This contained information on every issued address such as the final outcome, the screening outcomes, the observational data collected by interviewers, sample variables and geo-demographic variables.

Finally, a three-year data set was also supplied. This was a combined file containing data from the last three full survey years – from April 2017 – March 2020.

Data was delivered six weeks after the end of each quarterly fieldwork period. Each quarterly data delivery included interviews that were **achieved** in each specific 12-month period, rather than those that were **issued** in a specific time period. Thus, the four sets of quarterly data files delivered in 2019-20 covered all the relevant interviews achieved in the following periods:

- July 2018 – June 2019
- October 2018 – September 2019
- January 2019– December 2019
- April 2019 – March 2020³⁸

³⁸ The April 2019 – March 2020 data file is the data on which the 2019-20 annual crime figures are based and is the basis of the file deposited at the UK Data Archive.

8.3 Content of SPSS data file

The SPSS data files delivered to the Office for National Statistics contain various types of variables. The main types of variables contained on the files are:

- **Questionnaire variables** (NVF and VF).
- **Geo-demographic variables** (NVF only). All interviews had a set of pre-specified geo-demographic variables attached to them.
- **Observational variables** (NVF only). All interviews had the observational data collected by interviewers in the Electronic Contact Sheet attached to them (see Appendix C in Volume 2) These variables are included in the quarterly data files.
- **Coding variables** (NVF and VF). On the Non Victim File, 4-digit SOC2010 codes and NS-SEC classification are included for both the respondent and the Household Reference Person. On the Victim File, a full set of offence codes are attached as outlined in [Chapter 7](#).
- **Derived variables** (NVF and VF). Many derived variables are also added to the file. These consisted primarily of 2 types; flag variables and classificatory variables
 - **Flag variables** (NVF and VF) that identify, for example, the type of sample, the part-sample module split and sub-split, the date of interview, the month of issue, whether a partial or full interview, whether a victim or non-victim, etc. On the Victim File, flag variables include whether the record was a long or short victimisation module, whether it was a series or a single incident, and whether it was inside or outside the reference period.
 - **Classificatory variables** (NVF only) derived from the data. These included standard classifications such as ONS harmonised variables, banded age groups, ethnic groups, income groups, etc.
- **Weighting variables** (NVF only).

8.4 Case identifier

The case identifier is designed to meet the requirements of a continuous survey.

On the Non-Victim File, where each individual case or record represents an individual respondent, the unique case identifier (ROWLABEL) is an 8 or 9 digit number constructed as follows:

	Column position	Values
Year of issue	1-2	1-99
Area point number	3-6	1000-9999
Address number	7-9	1-99
Screen number ³⁹	9	0 or 8

On the Victim File, where each individual case or record represents a victimisation module, the unique case identifier (MATCH) is a 10-digit number, which is identical to ROWLABEL with the addition of the victimisation module number:

³⁹ Screen numbers are used to identify the type of sample. '0' indicates a core sample case and '8' indicates an interview with a 10 to 15 year old.

	Column position	Values
Year of issue	1-2	1-99
Area point number	3-6	1000-9999
Address number	7-8	1-99
Screen number	9	0 or 8
Victimisation module number	10	1-6

8.5 Naming conventions

In creating the 2019-20 data files great attention was paid to ensuring as much consistency as possible was maintained with previous years of the survey. Variable names were kept the same as on previous surveys wherever possible. Consistency is particularly important on a continuous survey where data from one survey year is combined with data from a previous survey year as described in section 8.2. However, this means it is also important to systematically document changes to questions over time to avoid confusion amongst users. For example, small changes to a question from one year to the next (such as adding an extra code to the code frame) can create the possibility of wrongly merging data that appears similar but, in fact, is not. To avoid such situations, the variable names on the 2019-20 data file were changed to reflect any variables where such changes had been introduced between 2018-19 and 2019-20 (see [Table 8.1](#)).

Table 8.1 Changes in variables between 2018-19 and 2019-20 survey

Module	2018-19 variable	2019-20 variable	Reason for change
Self-Completion Module: Drug Use And Drinking	drqwho2	drqwho3	Change to code frame
Self-Completion Module: Drug Use And Drinking	drqwho2o	drqwho3o	Change to code frame
Core Victim File			
Fraud Victim Form	fv86	fv862, fv86a, fv86b	Change to question wording
Fraud Victim Form	frspmona - k	frspmon2a - j	Change to code frame
Fraud Victim Form	faware2a - k	faware3a - j	Change to code frame
Fraud Victim Form	fawareA	fawarea2	Change to code frame
Fraud Victim Form	foffrel3	foffrel4	Routing into question changed

Table 8.2 Geo-demographic variables added to the survey in 2018-19

Variable	Comments
lad19cd	Added
lad19nm	Added
lad19ncd	Added
lad19typ	Added
atyp2019	Added
agrp2019	Added
acat2019	Added
emdidc19	Added
eincdc19	Added
eempdc19	Added
eedudc19	Added
eheadc19	Added
ecridc19	Added
ehoudc19	Added
eenvdc19	Added
edacdc19	Added
edopdc19	Added

8.6 Don't Know and Refused values

The convention for Don't Know and Refusal codes used in the most recent surveys was maintained on the 2019-20 data. This meant that on the SPSS file the code for Don't Know was '9' for code frames up to 7, '99' for code frames up to 97, and so on. The code for Refused was 8, 98, and so on. Since these are standard codes used throughout the SPSS files, Don't Know and Refused codes are not labelled.

8.7 Multiple response variables

Prior to the 2001 survey, multiple response variables were created as a set of variables equal to the maximum number of answers that could be given. The first variable held the first answer given by the respondent; the second variable held the second answer given, and so on.

After discussions with the Home Office it was agreed from 2001 onwards to present multiple response variables differently from previous years. Multiple response variables were set up as a set of variables equal to the total number of answers possible (including Don't Know and Refused). Each variable was then given a value of '0' or '1' depending on whether the respondent gave that particular answer or not. To denote this change, all multiple response variables in 2001 were all named with a letter suffix, rather than the number suffix that was used in previous years of the survey.

An example of a multiple response variable where there are seven possible answer categories, and so seven separate variables, is shown below:

AGEOFF2A-		
AGEOFF2G	[ASK IF NumOff=1]	
How old was the person who did it? Would you say [he/she] was...READ OUT		
CODE ALL THAT APPLY		
1.	a child aged under 10	(AGEOFF2A)
2.	a child aged between 10 and 15	(AGEOFF2B)
3.	aged between 16 and 24	(AGEOFF2C)
4.	aged between 25 and 39	(AGEOFF2D)
5.	or aged 40 or over?	(AGEOFF2E)
	Don't Know	(AGEOFF2F)
	Refused	(AGEOFF2G)

8.8 Data output on the 10 to 15 year olds survey

The data for the 10 to 15 year olds survey is delivered to ONS to the same quarterly timetable as the core survey data. As with the core data two data files are supplied, the Non Victim File and the Victim File.

9. Weighting

9.1 Overview of weighting

The following weights have been calculated for the 2019-20 CSEW data:

- A household weight for the core sample
- An individual adult weight for the core sample

In addition to these weights, the Office for National Statistics apply additional calibration weights once they receive the data so that the data reflect the population profile by age and gender within region ([see section 9.10](#)).

There are three main reasons for computing weights on the CSEW:

- To compensate for unequal selection probabilities. In the CSEW, different units of analysis (households, individuals, instances of victimisation) have different probabilities of inclusion in the sample due to factors such as over sampling of smaller police force areas, the selection of one dwelling unit at multi-household addresses, the selection of one adult in each household, and the inclusion of a single victimisation module to represent a series of similar incidents.
- To compensate for differential response. Differential response rates can arise both between different geographic units (e.g. differences in response between regions or between different types of neighbourhood) and between different age and gender sub-groups.
- To ensure that quarters are equally weighted for analyses that combine data from more than one quarter.

As outlined above a variety of different weights were computed to meet the different analysis requirements. The 2019-20 weighting schedule was the same as the weighting schedule applied on previous surveys.

All weights include a component to compensate for unequal selection probabilities, while components to compensate for differential response and to equally weight quarters are included in some weights but not in others.

9.2 Component weights

The weights constructed for the 2019-20 CSEW dataset were based on a number of components. The following conventions were used for the components that made up the final weights:

- **w₁**: weight to compensate for unequal address selection probabilities between police force areas;
- **w₂**: 'address non-response weight' to compensate for the observed variation in response rates between different types of neighbourhood;
- **w₃**: dwelling unit weight;
- **w₄**: individual selection weight to account for different sized households; and
- **numinc**: a weight applied based on the number of incidents in each series

9.3 Police Force Area weight (w_1)

Under the survey design introduced in 2012 the address sampling probability varies *between* police force areas but not within.

The police force area weight (w_1) is proportional to one divided by the address sampling probability.

9.4 Address non-response weight (w_2)

From April 2013, a new 'address non-response' weight replaced the 'inner city' weight as a method for compensating for variation in response rates between different types of area⁴⁰. Previously, each address was classified as 'inner city' or otherwise and a weight (w_2) given to responding cases from each class equivalent to one divided by the class response rate. Under the new method, responding cases are given a weight (w_2) equivalent to one divided by its estimated response probability.

This estimated response probability is calculated for each responding case based on four factors. These factors were selected following an analysis project carried out in 2012. The four factors are:

- 2011 Census Output Area Classification (twenty-one 'group' level)
- Region
- Proportion of households in local LSOA that contain only one person (Census 2011)
- ONS Urbanity indicator (twelve categories, updated based on Census 2011)

The estimated response probability of each responding case is derived from an analysis of the most recent twelve months of fieldwork assignments for which we have final outcome data for every address. A logistic regression model of response probability is fitted to this data to obtain a set of coefficients which can be applied to each responding case in the released dataset.

The advantage of this method over the previous 'inner city' weighting method is that a greater variety of factors are taken into account and the result should be a more accurate estimate of response probability for each case.

9.5 Dwelling unit weight (w_3)

At addresses which had more than one dwelling unit (defined as structurally separate properties which have their own lockable front door, or their own letter boxes, or their own bells but which share the same address), one dwelling unit was selected at random by a computer algorithm built into the electronic contact sheet. The dwelling unit weight is therefore simply the number of dwelling units identified at the address. In the vast majority of cases, the dwelling unit weight is 1. From 2014, this weight also includes a component to reflect any sampling of households within the sampled dwelling unit. This is a rare occurrence but w_3 is technically equal to the number of dwelling units at the address multiplied by the number of households in the sampled dwelling unit.

Weight w_3 is capped at 4 to limit the variance of these weights.

9.6 Individual weight (w_4)

At dwelling units that had more than one eligible adult, one adult was selected at random by a computer algorithm built into the electronic contact sheet. This means that the probability of any one individual being selected is inversely proportional to the number of adults in the household. The individual weight is therefore usually the number of adults in the household.

⁴⁰ Details of how the inner city weight was constructed can be found in the 2006/07 BCS technical report volume 1.

Weight w_4 is capped at 5 to limit the variance of these weights.

Furthermore, the product of the dwelling unit weight w_3 and the individual weight w_4 is capped at 5 for those weighted analyses that use both components.

In a small number of cases, the number of adults recorded during the doorstep screening process was different from that recorded in the subsequent interview. This was primarily due to either the interviewer being given wrong information by a household member or a change in the household composition between screening and interview. In such cases the interviewer was not required to re-do the selection process except under very specific circumstances. To ensure that the correct probability of selection is maintained the individual weight is always based on the number of adults recorded at the screening stage and not the number of adults recorded during the interview.

9.7 Series weight (numinc)

This weight is applied when estimating victimisation rates. For single incidents the weight is set to 1. For series incidents, where only details are collected about the most recent incident in the series, the weight equals the number of incidents in the series that fall within the reference period, subject to a maximum limit that is specific to the offence code group⁴¹. Table 9.1 shows the maximum limits used for 2019-20 data. These limits are equal to *either* (i) the 98th percentile series incident count over the period April 2016 to Mar 2019, *or* (ii) 5, whichever is the higher value.

⁴¹ Although the number of incidents is capped for weighting purposes, the actual number of reported incidents in each series (uncapped) is also supplied on the data file.

Table 9.1 Limits to 2019-20 series weights for each offence code group

Offence code group	Weight limit
INDIVIDUAL LEVEL OFFENCES	
Violence excepting sex offences, threats and robbery (codes 11,12,13,21,32,33)	10
Sex offences (codes 31,34,35)	5
Threats (codes 91,92,93,94)	9
Robbery (codes 41, 42)	5
Personal theft (codes 43,44,45)	5
Other personal theft (codes 67, 73)	5
Fraud (codes 200,201,202,203,204,205,206,207,208,210,211,212)	5
Computer misuse (codes 320,321,322,323,324)	5
HOUSEHOLD LEVEL OFFENCES	
Burglary (codes 50,51,52,53,57,58)	5
Other household theft (codes 55,56,65)	5
Motor vehicle crime (codes 60,61,62,63,71,72)	5
Bike theft (code 64)	5
Vandalism (codes 80,81,82,83,84,85,86)	5

In estimating victimisation levels, the household or individual weights are multiplied by the numinc weight, according to which offence classification code has been assigned to the incident(s).

9.8 Core sample weights

The main units of analysis used on the CSEW are households, individuals, and incidents of victimisation. Different weights are used depending upon the unit of analysis. In particular, some crimes are considered household crimes (e.g. burglary, vandalism to household property, theft of and from a car) and therefore the main unit of analysis is the household, while others are personal crimes (assault, robbery, sexual offences) and the main unit of analysis is the individual.

For the core sample two design weights are constructed to take account of this difference, namely the **core household weight** and the **core individual weight**. These are calculated as follows:

$$\mathbf{wtm2hhu} = W_1 * W_2 * W_3$$

$$\mathbf{wtm2inu} = W_1 * W_2 * W_3 * W_4$$

Note that both w_3 and w_4 are capped to avoid extreme values (see above). Although capping of extreme weights may introduce a small amount of bias this is more than compensated for by the improvement in precision that results. The capped weights are called **wtm2hhf** and **wtm2inf** respectively.

Finally, the weights are scaled to a notional sample size of 8,625 interviews per quarter. Although an approximately equal number of addresses are normally issued each quarter, the number of interviews actually achieved per quarter varies to some extent. For analyses based upon a 12 month period, the weights are constructed to adjust for differences in sample size by equalising the quarterly achieved sample sizes.

The final scaled weights are called **wtm2hhs** and **wtm2ins** respectively.

9.9 Weighting on the 10 to 15 survey

A logistic regression model is used to estimate the response probability of the selected 10-15 year old, *given* other data known about the child, the household and the sampled adult. The model was originally developed in 2009 but updated in 2015 and includes the parameters listed below. The coefficients applied to each parameter are updated on a biennial basis. The coefficients used for the 2019-20 survey were derived from data collected between January 2017 and December 2018 inclusive.

Parameters used to estimate response probability for each 10-15 year old:

- Age of sampled child
- Gender of sampled child
- Relationship of sampled child to interviewed adult
- Whether sampled child has own mobile phone
- Marital status of the household reference person (HRP)
- Adult respondent's opinion about the police

The final weight produced for each case in the 10-15 year old sample is equal to the household weight **wtm2hhs** multiplied by the product of (i) the reported number of 10-15 year olds in the household, and (ii) the estimated (conditional) response probability as derived from the logistic regression model described above. The product of component (i) and the dwelling unit component of **wtm2hhs** (w_3) is capped at 4 to prevent excessive variation in the design weights. Furthermore, to guard against errors due to model misspecification, the respondents are ranked by component (ii) and 'binned' into five equal-sized groups. The group mean response probability is used *in place of* the individual response probability when constructing the final weight.

This weight is then scaled so that each interview quarter has the same sum of weights (750) as each other.

9.10 Calibration Weights

Once the data is sent to ONS a further set of calibration weights are calculated and applied to counter the effect of differential response rates between age, gender and regional sub-groups. Results for CSEW surveys from 1992 onwards have all been re-weighted using this technique⁴².

The calibration weighting is designed to make adjustments for known differences in response rates between different age and gender sub-groups and for households with different age and gender composition. For example, a 24 year old male living alone may be less likely to respond to the survey than one living with a partner and a child. The procedure therefore gives different weights to different household types based on

⁴² Calibration weights are applied to the data by ONS after the application of the base weights computed by Kantar.

their age and gender composition in such a way that the weighted distribution of individuals in the responding households matches the known distribution in the population as a whole.

The effects of applying these weights are generally low for household crime, but are more important for estimates of personal crime, where young respondents generally have much higher crime victimisation rates than average, but also lower response rates to the survey. However, crime trends since the 1992 survey have not been altered to any great extent by the application of calibration weights. The calibrated weight variables are **c11hhdwgt** (households) , **c11indivwgt** (individuals aged 16+), **c11cindivwgt** (individuals aged 10-15) and **c11weighti** (incidents to households or individuals aged 16+).

10. Comparing key survey variables with the population

In order to assess the representativeness of the final achieved sample this chapter compares the profile of the 2019-20 survey against population estimates for a range of socio-demographic variables. In addition to comparing the age and sex profile of the survey with the latest population estimates comparisons are also made with data from the 2011 Census.

The tables presented below show the survey profile with the appropriate design weights applied (either household or individual weight) but without the application of the calibration weighting. Comparisons are made based on the 2019-20 achieved sample (i.e. from April 2019 to March 2020) rather than on the 2019-20 issued sample.

10.1 Regional distribution of the sample

Table 10.1 shows the distribution of households by region in the 2019-20 survey compared with the 2011 Census⁴³. This shows that the regional profile of the weighted sample was broadly in line with the population distribution.

Table 10.1 Distribution of households by region in the 2019-20 survey compared with the 2011 Census

	2019-20 CSEW	2011 Census	Difference
	%	%	%
North East	6.1	4.8	1.3
North West	12.7	12.9	-0.2
Yorkshire and The Humber	9.2	9.5	-0.3
East Midlands	9.0	8.1	0.9
West Midlands	9.0	9.8	-0.8
East of England	11.2	10.4	0.8
London	11.4	14.0	-2.6
South East	13.5	15.2	-1.7
South West	10.3	9.7	0.6
Wales	7.4	5.6	1.8

⁴³ All Census figures presented in the tables are sourced from <http://www.nomisweb.co.uk/census/2011>

10.2 Age and sex profile of the sample

Table 10.2 shows a comparison between the achieved 2019-20 core adult sample and the mid-2019 population estimates for England and Wales by sex and age. This shows that the survey slightly underrepresented men and those aged under 35. The profile of the survey by sex and age was similar to previous years. These patterns are fairly typical of large-scale surveys and reflect the lower co-response rates generally achieved among these particular groups.

Table 10.2 Age and sex profile of adult sample against mid-2019 population estimates

	2019-20 CSEW	Mid-2019 population estimates	Difference
	%	%	%
Sex			
Male	46.1	49.0	-2.9
Female	53.9	51.0	2.9
Men			
16-19	2.4	5.7	-3.3
20-24	4.0	8.1	-4.1
25-34	13.4	17.1	-3.7
35-44	15.5	15.8	-0.3
45-54	16.3	16.9	-0.6
55-64	17.4	15.0	2.4
65-74	16.9	12.1	4.8
75-84	11.2	6.9	4.3
85 and over	3.0	2.3	1.3
Women			
16-19	1.9	5.2	-3.3
20-24	3.9	7.3	-3.4
25-34	14.9	16.2	-1.3
35-44	16.7	15.4	1.3
45-54	16.1	16.6	-0.5
55-64	16.0	14.9	1.1
65-74	15.6	12.6	3.0
75-84	11.3	8.0	3.3
85 and over	3.6	3.8	-0.2

Table 10.3 shows a similar comparison for the 2019-20 10-15 year olds survey. This shows that the survey slightly under represented girls (particularly those aged 14 years old) and boys aged 10 and 11 years old.

Table 10.3 Age and sex profile of 10 to 15 year olds sample against mid-2019 population estimates

	2019-20 CSEW	Mid-2019 population estimates	Difference
	%	%	%
Sex			
Boys	51.5	51.3	0.2
Girls	48.5	48.7	-0.2
Boys			
10	15.8	17.3	-1.5
11	15.5	17.5	-2.0
12	17.9	17.0	0.9
13	16.7	16.6	0.1
14	17.1	15.9	1.2
15	17.0	15.7	1.3
Girls			
10	16.5	17.3	-0.8
11	17.3	17.5	-0.2
12	19.4	16.9	2.5
13	16.3	16.6	-0.3
14	14.3	16.0	-1.7
15	16.1	15.7	0.4

Although not reported here, as already mentioned the age and sex distribution of the achieved sample is further corrected by ONS at the analysis stage through the application of calibration weights so that the age and sex profile of survey respondents match population estimates within each region (see [section 9.10](#)).

10.3 Other household characteristics

Table 10.4 shows the profile of the 2019-20 survey compared with some key household characteristics from the 2011 Census. This shows that the survey underrepresented three person households and larger households, which is probably related to the under representation of younger people seen above. Although housing tenure was broadly in line with the Census there was a noticeable under representation of people living in flats. This is almost certainly due to the lower response rate achieved at flats caused by the practical difficulties of negotiating access through entry phone systems. Those who do not own a car or van are also slightly underrepresented.

Table 10.4 Household characteristic of the core adult sample against 2011 Census

	2019-20 CSEW	2011 Census	Difference
	%	%	%
Tenure			
Owned	65.5	64.3	1.2
Social renting	16.5	17.5	-1.0
Private renting	18.0	18.2	-0.2
Accommodation type			
Whole house or bungalow	85.0	78.6	6.4
Flat, maisonette or apartment	14.7	20.7	-6.0
Household size			
1 person household	28.5	30.2	-1.7
2 person household	37.2	34.2	3.0
3 person household	15.4	15.6	-0.2
4 or more person household	18.9	19.9	-1.0
Car ownership			
No cars or vans	20.2	25.6	-5.4
1 car or van	42.0	42.2	-0.2
2+ cars or vans	37.8	32.1	5.7

10.4 Other individual characteristics

Table 10.5 shows the profile of the 2019-20 survey compared with some key individual characteristics from the 2011 Census. Again, the profile of the survey is broadly in line with the Census across all dimensions. There is a slight under representation of those who have never worked or are long term unemployed and those in intermediate occupations. There is also an over representation of those who work in higher managerial, administrative and professional occupations and those who report having no religion. This is largely caused by gradual increases in the proportion who report having no religion over the last 5 years.

Table 10.5 Comparison of individual respondent characteristic against 2011 Census

	2019-20 CSEW	2011 Census	Difference
	%	%	%
NS-SEC⁴⁴			
Higher managerial, administrative and professional occupations	38.3	34.2	4.1
Intermediate occupations	23.7	24.4	-0.7
Routine and manual occupations	34.5	35.3	-0.8
Never worked and long-term unemployed	3.5	6.2	-2.7
Ethnic group			
White	89.1	88.2	0.9
Mixed/multiple ethnic group	1.2	1.2	0.0
Asian/Asian British	6.1	6.9	-0.8
Black/African/Caribbean/Black British	2.8	2.9	-0.1
Other ethnic group	0.8	0.8	0.0
Religion			
No religion	37.0	25.8	11.2
Christian	55.8	66.0	-10.2
Buddhist	0.4	0.5	-0.1
Hindu	1.3	1.6	-0.3
Jewish	0.4	0.5	-0.1
Muslim	3.6	4.3	-0.7
Sikh	0.5	0.8	-0.3
Other	0.5	0.5	0.0

⁴⁴ 16-74 year olds only